Applicability of

FACTOR ANALYSIS IN THE BEHAVIORAL SCIENCES

A methodological study

by

STEN HENRYSSON

BUREAU OF EDNL. & PSYL. RESEARCH
LIBRARY.

S.C.E.R.T., West Bengal
Library
Calcutta
25/3, B C. Road

ALMQVIST & WIKSELL
STOCKHOLM

C.E R T., West Bengal
ate....16-2-59
oc. No....1309

311
HEN

UPPSALA 1957
ALMQVIST & WIKSELLS BOKTRYCKERI AB

PREFACE

The purpose of this monograph is to determine the kind of problems occurring in the behavioral sciences, which can be solved either completely or partly by factor analysis, and how this can best be accomplished.

As a background to our study, we shall first discuss the characteristics of the factor analytical models and the assumptions on which they are based. That factor analysis has a wide field of application as a descriptive condensing aid would appear to be indisputable. There are differencies of opinion, however, about the possibility of finding factors which can be included in explanatory models. This is partly due to the fact that the meaning of explanatory factor analysis is not clear. Consequently an attempt will be made to find a meaning for this kind of analysis. We shall discuss, from this point of view, the question of how to conduct an explanatory factor analysis, particularly when the principles of simple structure are applied.

The theory of factor analysis embraces a number of comprehensive and complex problems, e.g., of a statistical nature. Some of these problems still require to be solved. This monograph will, however, be concerned with the questions which are essential to the main problem, and will disregard a number of important but secondary problems. The monograph is not intended to cover the discussion of the whole field of factor analysis. Nor will the comments and criticisms, which have already appeared and been refuted in psychological and statistical literature, be considered here.

The monograph mainly deals with questions of current interest in order to clarify them, to discuss the problems yet unsolved, to describe the attempts that have been made to solve them, and to indicate along which lines the solutions are likely to be found.

The problems dealt with are related to various sciences, such as mathematics, statistics, psychology, and philosophy. From the viewpoint of each of these branches of science, this account is perhaps incomplete and imperfect and cannot claim to be "watertight". An attempt is made to integrate the results obtained in different special sciences by investigating how factor analytical methods can be used to solve problems in the behavioral sciences. In any case a useful purpose may be served if someone, who is fairly well acquainted with the sciences referred to above, can formulate problems and suggest solutions in such a way, that scientists, who specialize in a more limited field, will obtain a better position from which they can attack some of these problems.

It is obviously impossible, within the scope of this monograph, to give a complete account of factor analytical theory. Only the background which is essential for the discussion can be given. For a fuller account, the reader is referred to classical handbooks such as Thurstone's *Multiple Factor Analysis* or Thomson's *Factorial Analysis of Human Ability* or Holzinger & Harman, *Factor Analysis*.

It has been found expedient to devote the greater part of this monograph to a purely mathematical treatment of factor analytical theory, and not to make any attempt to go fully into questions connected with sampling fluctuations. This aspect of the problem is discussed in the latter part of this study. This does not at all imply that the statistical viewpoint is of less importance, but is due to heuristical reasons. The monograph will be clearer and more readily understandable if the mathematical theory and its significance are presented at the outset without the complications which are inevitable, if sampling fluctuations are included. It should be pointed out however, that since the main aim of this monograph is to consider how factor analysis can solve problems in the behavioral sciences, the treatment of the purely statistical discussion is limited to what is necessary in order to fulfil this purpose.

During the time—since 1948—that I have devoted myself to the study of factor analytical problems, personal contacts have been established with a number of scientists working in this field. I have worked under Thomson in Edinburgh, Thurstone in Chicago, and Wold in Uppsala. When I was abroad, and also on other occasions, I have had the opportunity of personal discussions with Ahmavaara, Bargmann, M. S. Bartlett, Burt, Cronbach, Eysenck, J. W. French, Holzinger, Lawley, Lubin, Peel, Saunder, Tucker, Vernon, Whittle and other investigators. During our discussions some of these scientists have expressed ideas and points of view, which are not always to be found in their published works.

Professors Husén, Malmquist and Wold have read different sections of the manuscript and made suggestions and comments. It should be pointed out, however, that I am personally responsible for the statements made and the results given.

From *Statens Psykologisk-Pedagogiska Institut*, *Statens Samhälls- och Rättsvetenskapliga forskningsråd*, and *Sverige-Amerika Stiftelsen* I have received grants for travelling, and for assistance with calculation, typing and translation.

Many persons and institutions—besides those mentioned above—have helped me in various ways in the writing of this monograph. They are so numerous that it is impossible to name them all here. The gratitude I feel towards all of them—from my daughters Mari and Kersti to my teacher and friend Louis Leon Thurstone—is none the less great.

Stockholm, March 1957

Sten Henrysson

CONTENTS

PREFACE 5
CONTENTS 9
SYMBOLS AND TERMINOLOGY 11

I. THE BASIC THEORY OF FACTOR ANALYSIS

The Rise of Factor Analysis 13
The Basic Equations. 14
Component Analysis and Factor Analysis 22
The Geometrical Theory and Rotation 26
Assumptions about the Nature of the Data 29

II. THE MOST IMPORTANT FACTOR MODELS

The Need for Restrictive Assumptions 32
The Forerunners of Factor Analysis 32
The Spearman School 33
Burt's Factor Analytical Methods 40
Thomson's Contribution 43
The Thurstone School 44

III. DIFFERENT APPLICATIONS OF FACTOR ANALYSIS

The Problem of Finding the Number of Dimensions 50
The Problem of Finding Descriptive Factors 53
Descriptive Factor Analysis and Its Applications. 55
Prediction and Factor Analysis 59
Differential Prediction and Factor Analysis 68
Conclusions about Factor Tests as Predictors 73
The Problem of Finding Explanatory Factors 76
IV. FACTOR ANALYSIS AND THEORY OF SCIENCE 77
Description and Explanation in the Theory of Science 78
Intervening Variables and Hypothetical Constructs. 82
The Distinction between Descriptive and Explanatory Factor Analysis 86

V. Explanatory Factor Analysis

The Testing and the Generating of Hypotheses 89
Hypothesis-Generating Factor Analysis and the Principle of Parsimony 91

VI. Factor Analysis According to the Principles of Simple Structure

The Meaning of Simple Structure 95
Oblique Structure 100
The Selection of Test Battery 101
Communality Estimates 104
Analytical Methods for Rotation to Simple Structure 108

VII. The Problem of Invariance

The Concept of Invariance 111
Methods for Finding Invariance 115
Simultaneous Rotation to Invariant Solutions 118
Conclusions about the Methods of Finding Invariance 121
Nonfactorial Validation 122

VIII. Factor Analysis from a Statistical Point of View

The Need of Statistically Acceptable Methods 124
Maximum Likelihood Solutions and Least Squares Solutions . . 127
Tests of Significance for Factor Loadings 137
Summary 139
Appendix 143
References 150

SYMBOLS AND TERMINOLOGY

Although certain sections require a special set of symbols, generally the following notations and terminology are used in this monograph.

X_{ji}	raw score in test j for individual i
i	subscript for individual
j, t	two subscripts designating tests
N	total number of persons in a group
M	mean
s	standard deviation
z_{ji}	standard score in test j for individual i
r_{jt}	correlation between test j and t
F_{ki}	common factor score in the kth common factor for person i
S_{ji}	specific factor score in test j for person i
ε_{ji}	error factor score in test j for person i
U_{ji}	unique factor score in test j for person i
a_{jk}	loading of factor k in test j
h_j^2	communality of the jth test
Z	matrix containing standard scores
R_1	complete correlation matrix with unities in the diagonal
A	matrix containing all loadings
F	matrix containing all factor scores
R	reduced correlation matrix with communalities in the diagonal
A_m	matrix containing common factor loadings
F_m	matrix containing common factor scores
R_S	diagonal matrix containing specific parts of test variances
A_S	diagonal matrix containing specific loadings
F_S	matrix containing specific factor scores
R_ε	diagonal matrix containing error parts of test variances
A_ε	diagonal matrix containing error factor loadings
F_ε	matrix containing error factor scores
R_U	diagonal matrix containing unique parts of test variances
A_U	diagonal matrix containing unique factor loadings
F_U	matrix containing unique factor scores
Φ	matrix containing correlations between all factors
ϕ	matrix containing correlations between common factors
H	diagonal matrix containing communalities
A_0	matrix containing common factor loadings before rotation
Λ	transformation matrix
h_j, h_t	length of test vectors j and t
ϕ_{jt}	angle between test vectors j and t

CHAPTER I

THE BASIC THEORY OF FACTOR ANALYSIS

The Rise of Factor Analysis

Factor analysis has its roots in the psychological and biological discussions around Spencer, Galton, McDougall et al., during the later part of the nineteenth century. It was Spearman, however, who started the growth of factor analysis through a paper in 1904 on the nature of intelligence. In this paper he described a factor analytical investigation of various cognitive tests. He started out from a psychological theory which he tried to verify with the help of a factor analytical method. The basic idea in this method was to trace back the relations between suitably chosen cognitive variables to a general ability factor—the so-called *G*-factor. Each observed variable was also assumed to measure a specific factor not to be found in any of the other observed variables. For some decades discussion revolved principally around the validity of this simple model.

Through Thurstone particularly, development during the 1930's went beyond Spearman's limited model. More general models were developed and, at present, research efforts are mostly directed upon these models.

From the very beginning of factor analysis, an intensive debate has been going on about its basis and usefulness. This debate is still continuing and is characterized by the fact that the various factor analysts have different views as to the value of the methods. Even the same investigator's standpoint may vary. The present methodological development is characterized by working with different special problems connected with the various methods or with the general theories, and by seeking new fields of application. Since Thurstone's book, *Multiple Factor Analysis* (1947), and the third edition of Thomson's book, *Factorial Ana-*

lysis of Human Ability (1948), were published, no general texts, more important from a methodical viewpoint, have come out. Thus, Vernon's *The Structure of Human Abilities* (1950 a) is mainly of interest as to the application of the factor analysis. Spearman-Jones's book *Human Ability* (1950) added scarcely anything new to the discussions in Spearman's earlier work. Cattell's *Factor Analysis* (1952) and Fruchter's *Introduction to Factor Analysis* (1954) include accounts of the factor analytical methods and the technical procedures.

A growing interest in factor analysis has lately been noticed among statisticians. However, they have concentrated themselves mainly on statistically efficient but unrotated solutions, which need to be adapted and further developed in order to be really useful in the behavioral sciences.

The Basic Equations

Factor analysis supplies methods for reducing a large number of observed variables to a lesser number, in some way more fundamental variables or, as they are usually called, factors. This is usually done through an analysis of intercorrelations between the observed variables.

In this section of the monograph an account is given of the mathematical theory on which factor analysis is based. This account is not intended to be complete but will only bring out such characteristics of the factor analysis as are of importance in the following discussion. A more complete account is given, for example, by Holzinger-Harman (1941) or by Thurstone (1947). Elmgren (1955), who introduced factor analysis in Sweden, has published a textbook in Swedish on the subject.

The present monograph is concentrated on analyses of correlations between observed variables applied on large samples of individuals at a certain time, that is, on the so-called *R*-technique, which, at the present time, is the most developed and used technique. It would carry too far to go into various other problems, such as analysis of correlations between persons. Many of the principal questions regarding *R*-technique are, to a greater or lesser degree, of importance also for the other techniques. Further details about different techniques have been given by Cattell (1952, pp. 88–126). The discussion here is also limited to continuous variables having linear relations with each other. The treatment of discontinuous and non-linear cases is discussed, for example, by Coombs (1952), Lazarsfeld (1950), and Burt (1953).

This chapter contains the mathematical theory. The purely statistical discussion concerning factor analysis as a problem of estimation and significance has, as mentioned in the preface, been reserved to chapter VIII.

The basis for the theory are the different observations X_{ji} for N individuals in n different observed variables. For convenience the word "score" is used in the following instead of observation and "test" instead of observed variable. This, of course, does not imply that the methods can only be used on test scores.

Each one of the scores is expressed as the standard score z_{ji} in each test according to the following formula

$$z_{ji}=\frac{X_{ji}-M_j}{s_j} \quad (j=1, 2 \ldots n;\ i=1, 2, \ldots N). \tag{1}$$

M_j and s_j are the mean and standard deviation for the jth test. That means that

$$M_j=\frac{1}{N}\sum_{i=1}^{N} z_{ij}=0 \ \text{ and } \ s_j^2=\frac{1}{N}\sum_{i=1}^{N} z_{ji}^2=1. \tag{2}$$

The test scores expressed in standard scores for the N different individuals in the n tests can be shown in a test score matrix like the following

$$Z=\begin{bmatrix} z_{11} & z_{12} & z_{13} & \cdots & z_{1N} \\ z_{21} & z_{22} & z_{23} & \cdots & z_{2N} \\ z_{31} & z_{32} & z_{33} & \cdots & z_{3N} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ z_{n1} & z_{n2} & z_{n3} & \cdots & z_{nN} \end{bmatrix} \tag{3}$$

Since the test scores are expressed in standard scores, the correlation between the tests j and t for the N individuals according to the formula for the product moment correlation will be

$$r_{jt} = \frac{\sum_{i=1}^{N} z_{ji} z_{ti}}{N}. \tag{4}$$

All the intercorrelations between the different variables can be included in the following correlation matrix:

$$R_1 = \begin{bmatrix} r_{11} & r_{12} & r_{13} & \dots & r_{1n} \\ r_{21} & r_{22} & r_{23} & \dots & r_{2n} \\ r_{31} & r_{32} & r_{33} & \dots & r_{3a} \\ . & . & . & \dots & . \\ . & . & . & \dots & . \\ . & . & . & \dots & . \\ r_{n1} & r_{n2} & r_{n3} & \dots & r_{nn} \end{bmatrix} \tag{5}$$

The matrix algebraic expression for the product moment correlations will then be

$$R_1 = \frac{1}{N} Z Z'. \tag{6}$$

The aim of factor analysis is, expressed in mathematical language, to describe the results in the tests as functions of a few variables, which from certain standpoints are more "fundamental" or convenient than the tests. These variables searched for, are called factors and can be indicated by F_1, F_2, F_3, ... F_q. Each one of the different test scores z_j is thus regarded as a function of the factors according to the general formula $z_j = f(F_1 F_2 F_3 \dots F_q)$.

Factor analysis has hitherto practically only dealt with the case, when z_j is a linear combination of factors as

$$z_{ji} = a_{j1} F_{1i} + a_{j2} F_{2i} + \dots + a_{jk} F_{ki} + \dots + a_{jq} F_{qi}. \tag{7}$$

F_{ki} is the factor score of individual i in the kth factor, and a_{jk} is the kth factor loading in test j. Formula (7) can be regarded as a regression equation where z_j is the dependent variable, whereas the various F_{ki} and a_{jk} refer to q independent variables and their respective regression weights.

The fully developed system of factor equations for n tests and N individuals in expressed by

$$Z = AF. \tag{8}$$

The matrix A includes the loadings of the q factors in the n tests.

$$A = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1k} & \ldots & a_{1q} \\ a_{21} & a_{22} & \ldots & a_{2k} & \ldots & a_{2q} \\ . & . & \ldots & . & \ldots & . \\ . & . & \ldots & . & \ldots & . \\ a_{j1} & a_{j2} & \ldots & a_{jk} & \ldots & a_{jq} \\ . & . & \ldots & . & \ldots & . \\ . & . & \ldots & . & \ldots & . \\ a_{n1} & a_{n2} & \ldots & a_{nk} & \ldots & a_{nq} \end{bmatrix} \tag{9}$$

The matrix F includes the N individuals' scores in the q factors.

$$F = \begin{bmatrix} F_{11} & F_{12} & \ldots & F_{1i} & \ldots & F_{1N} \\ F_{21} & F_{22} & \ldots & F_{2i} & \ldots & F_{2N} \\ . & . & \ldots & . & \ldots & . \\ . & . & \ldots & . & \ldots & . \\ F_{k1} & F_{k2} & \ldots & F_{ki} & \ldots & F_{kN} \\ . & . & \ldots & . & \ldots & . \\ . & . & \ldots & . & \ldots & . \\ F_{q1} & F_{q2} & \ldots & F_{qi} & \ldots & F_{qN} \end{bmatrix} \tag{10}$$

The main interest in factor analysis is at present concentrated on how to find methods for calculating loadings in matrix A.

The variance of a test j in terms of factors is found through the variance formula applied on equation (7).

$$\left.\begin{aligned} s_j^2 &= \frac{\sum\limits_{i=1}^{N} z_{ji}^2}{N} = \\ &= a_{j1}^2 \frac{\sum\limits_{i=1}^{N} F_{1i}^2}{N} + a_{j2}^2 \frac{\sum\limits_{i=1}^{N} F_{2i}^2}{N} + \cdots + a_{jk}^2 \frac{\sum\limits_{i=1}^{N} F_{ki}^2}{N} + \cdots + a_{jq}^2 \frac{\sum\limits_{i=1}^{N} F_{qi}^2}{N} + \\ &\qquad + 2\left(a_{j1}\, a_{j2} \frac{\sum\limits_{i=1}^{N} F_{1i}\, F_{2i}}{N} + \cdots + a_{j(q-1)}\, a_{jq} \frac{\sum\limits_{i=1}^{N} F_{(q-1)i}\, F_{qi}}{N}\right) \\ &= 1. \end{aligned}\right\} \tag{11}$$

All the factor scores are for convenience assumed to be in standard form and consequently all test variances are equal to one.

Formula (11) can thus be written

$$\begin{aligned} s_j^2 = a_{j1}^2 + a_{j2}^2 + \cdots + a_{jk}^2 + \cdots + a_{jq}^2 + \\ + 2\left(a_{j1}\, a_{j2}\, r_{F_1 F_2} + \cdots + a_{j(q-1)}\, a_{jq}\, r_{F_{(q-1)} F_q}\right) = 1. \end{aligned} \tag{12}$$

If the factors are uncorrelated, the section within the parenthesis is dropped. The result will then be

$$s_j^2 = \sum_{k=1}^{q} a_{jk}^2 = 1. \tag{13}$$

The determination of the factor loadings in the matrix A is made by the use of the information about the test that is found in the correlation matrix

$$R_1 = \frac{1}{N} Z Z'. \tag{14}$$

Since $Z = AF$, the relation will be

$$R_1 = \frac{1}{N} AF(AF)' = \frac{1}{N} AFF' A'. \tag{15}$$

In this relation the matrix

$$\frac{1}{N} F F' = \Phi \tag{16}$$

will include the correlations between the factors.

If the factors are uncorrelated, Φ is $= I$, i.e. a unit matrix. In that case the formula (15) is reduced to

$$R_1 = A A'. \tag{17}$$

The formula for a single correlation r_{jt} is

$$\begin{aligned} r_{jt} = a_{j1} a_{t1} + a_{j2} a_{t2} + \cdots + a_{jk} a_{tk} + \cdots + \\ + a_{jQ} a_{tQ} + (a_{j1} a_{t2} + a_{t1} a_{j2}) r_{F_1 F_2} + \cdots + \\ + (a_{j(Q-1)} a_{tQ} + a_{t(Q-1)} a_{jQ}) r_{F_{(Q-1)} F_Q}. \end{aligned} \tag{18}$$

If the factors are uncorrelated the single correlation r_{jk} is

$$r_{jk} = \sum_{k=1}^{Q} a_{jk} a_{tk}. \tag{19}$$

The meaning of this is that the test results are considered to be linear functions of a number of factors according to certain regression equations. The regression coefficients—loadings—in these equations must be selected in accordance with the correlations between the tests so that the relation $R_1 = A \Phi A'$ is fulfilled.

The factor analytical theory given above has treated the case when R_1 is a correlation matrix. With somewhat different matrix algebraic formulae, a similar theory can be applied to covariance matrices.

The main problem in factor analysis is to find the loadings a_{jk} in the matrix A which fulfil the condition $R_1 = A \Phi A'$. When A and Φ are found, then F can be estimated. Nothing prevents F from being calculated before A, or both being calculated at the same time. At the present stage most scientists consider that A is of greater interest than F since A indicates *which* factors and how much of each factor, on the average, the test measures (Lawley, 1953, p. 35). Estimates of the factor values F_{ki} for each individual would first seem to be of interest at a later stage when a number of, in some sense, fundamental factors have been established. The main problem is at present, therefore, to find the loadings from the intercorrelations between the tests, i.e. to solve A in the relation $R_1 = A \, \Phi \, A'$.

Unfortunately, in the general case, the correlation matrix R_1 does not include a sufficient number of restrictive conditions for a unique solution. Therefore further restrictions must be included. In addition, R_1 is in practice impaired by sampling errors and errors of measurement which complicate the estimation.

A principle common to most factor analytical methods is that the number of factors should be small, usually equal to the rank r of the correlation matrix. The addition of this condition, however, is not sufficient to obtain a unique solution.

Two major types of factors are distinguished, namely, *common* and *unique* factors. The first type refers to such factors as are loaded in at least two of the tests in the correlation matrix. The common factors therefore correspond to those parts of the variances that represent the intercorrelations between the various tests. The other major type—the unique factors—corresponds to that part of the test variance which is not common to this and any other test in the battery of tests. The unique factors are therefore uncorrelated with all other tests included in the battery. The unique factors are relative to their scope. By changing the composition of the battery of tests, one can displace the content of the unique factors in each one of the tests. If a new test is added which has a factor, which was earlier unique, common with another test, then this unique factor is changed into a common factor. In the same way, the removal of a test from a battery can result in a common factor becoming unique.

A unique factor can be divided into a specific part and an error part. The later corresponds to the unreliability of the test. The specific part is consequently equal to the reliable part of the unique variance.

Divided into three types of factors, the factor equation for the *i*th individual in the *j*th test gets the following expression:

$$z_{ji} = \sum_{k=1}^{m} a_{jk} F_{ki} + a_{jS} S_{ji} + a_{j\varepsilon} \varepsilon_{ji}. \tag{20}$$

F_{ki} designates common factors, S_{ji} the specific factor, and ε_{ji} the error factor. The matrix algebraic formulation of the above equations is

$$Z = A_m F_m + A_S F_S + A_\varepsilon F_\varepsilon. \tag{21}$$

A_m includes all the common factors, A_S the special ones, and A_ε the error factors. The matrices with the factor loadings are as follows:

$$A_m = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1k} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2k} & \dots & a_{2m} \\ \cdot & \cdot & \dots & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot & \dots & \cdot \\ a_{j1} & a_{j2} & \dots & a_{jk} & \dots & a_{jm} \\ \cdot & \cdot & \dots & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot & \dots & \cdot \\ a_{n1} & a_{n2} & \dots & a_{nk} & \dots & a_{nm} \end{bmatrix} \tag{22}$$

$$A_S = \begin{bmatrix} a_{1S} & & & & & & & \\ & a_{2S} & & & & & & \\ & & \cdot & & & & & \\ & & & \cdot & & & & \\ & & & & a_{jS} & & & \\ & & & & & \cdot & & \\ & & & & & & \cdot & \\ & & & & & & & \cdot \\ & & & & & & & & a_{nS} \end{bmatrix} \tag{23}$$

$$A_\varepsilon = \begin{bmatrix} a_{1\varepsilon} & & & & & & & \\ & a_{2\varepsilon} & & & & & & \\ & & \cdot & & & & & \\ & & & \cdot & & & & \\ & & & & a_{j\varepsilon} & & & \\ & & & & & \cdot & & \\ & & & & & & \cdot & \\ & & & & & & & \cdot \\ & & & & & & & & a_{n\varepsilon} \end{bmatrix} \tag{24}$$

The two later matrices are diagonal matrices which follow the formulae

S.C.E.R.T., West Bengal
Date 16-2-59
Acc. No. 1309

S.C.E.R.T., West Bengal
Library

$$A_S A_S' = A_S^2 \tag{25}$$

and
$$A_\varepsilon A_\varepsilon' = A_\varepsilon^2. \tag{26}$$

The matrix $R_1 = A \Phi A'$ is composed of the following three correlation matrices, namely

$$R = A_m \phi A_m', \tag{27}$$

$$R_S = A_S^2, \tag{28}$$

$$R_\varepsilon = A_\varepsilon^2. \tag{29}$$

R contains those parts of the correlations referring to common factors. The matrix ϕ contains the correlations between the common factors and is of order m. Its relation to Φ is

$$\Phi = \begin{bmatrix} \phi & 0 \\ 0 & I \end{bmatrix}. \tag{30}$$

I is here a unit matrix of order n. The matrix R_S is a diagonal matrix which includes the specific parts of the test variances while R_ε includes the error variances. The correlation matrix R_1 can be written

$$R_1 = R + R_S + R_\varepsilon = A_m \phi A_m' + A_S^2 + A_\varepsilon^2. \tag{31}$$

The unique parts of the variances are in the diagonal matrix

$$R_U = A_U^2 = R_S + R_\varepsilon = A_S^2 + A_\varepsilon^2. \tag{32}$$

This implies that the relation between unique factor loadings on the one hand and specific and error factors on the other, is the following:

$$A_U = (A_S^2 + A_\varepsilon^2)^{1/2}. \tag{33}$$

Component Analysis and Factor Analysis

In conjunction with the division in common and unique factors one can make a distinction used, among others, by Kendall (1950) and Rao (1955). They distinguish between component analysis and factor analysis. In component analysis one does not maintain the division between common and specific factors and analyzes the whole correlation matrix R_1 in what is called components, by using ones in the diagonal, i.e., the whole test variances.

In this case, the mathematical model is

$$\left.\begin{array}{l} z_{ji} = a_{j1} F_{1i} + a_{j2} F_{2i} + \cdots + a_{jk} F_{ki} + \cdots + a_{jq} F_{qi} \\ \quad (j = 1, 2 \ldots n;\; i = 1, 2 \ldots N;\; k = 1, 2 \ldots q). \end{array}\right\} \tag{34}$$

The aim of component analysis is a transformation of the n number of tests to q number of components that have certain suitable characteristics. The components will in this case give the same exact descriptions of the individuals as the tests do. The individual test results can be obtained as weighted sums of component scores and conversely.

The component loadings a_{jk} in the matrix A are calculated from the relation

$$R_1 = A \, \Phi \, A'. \tag{35}$$

The most common solution is the principal axis solution. It traces R_1 back to uncorrelated components ($\Phi = I$), which have the characteristic that the first of them absorbs as much of the variances as possible, the next one as much as possible of the rest, and so on. In this way a considerably lesser number of components than the number of tests can often express the main part of the variances of R_1 and make possible estimation of test scores close to their correct values. However, in psychological applications, one can as a rule count on the fact that the rank in the matrix, R_1 is equal to n and that thus $q = n$ components are required for a complete representation of the total variances in all the tests.

Another case of component analysis is an analysis of the correlation matrix R_1, where the variances of the diagonal are divided into reliable and unreliable parts, the later corresponding to errors of measurement. In this case the mathematical model is

$$z_{ji} = a_{j1} F_{1i} + a_{j2} F_{2i} + \ldots + a_{jk} F_{ki} + \ldots + a_{jq} F_{qi} + a_{j\varepsilon} \varepsilon_{ji}. \tag{36}$$

The loadings in this case follow the relation

$$R_1 = R_2 + R_\varepsilon = A_2 \, \Phi \, A_2' + A_\varepsilon^2. \tag{37}$$

In the matrix R_1, the reliable part corresponds to the matrix $R_2 = A_2 \Phi A_2$, and the unreliable part to $R_\varepsilon = A_\varepsilon^2$. The number of loadings to be estimated is $q + n$. As previously mentioned, however, the number of common factors in psychological applications are, in practice, equal to n. The total number of factors will then be $2n$.

The other type of analysis, factor analysis, includes specific factors as well. The mathematical model is

$$z_{ji} = \sum_{k=1}^{m} a_{jk} F_{ki} + a_{jS} S_{ji} \tag{38}$$

or

$$z_{ji} = \sum_{k=1}^{m} a_{jk} F_{ki} + a_{jS} S_{ji} + a_{j\varepsilon} \varepsilon_{ji}. \tag{39}$$

The model one selects depends on whether error factors are considered or not. Accordingly, the analysis works with correlation matrices of the type $R_1 = R + R_S$ or $R_1 = R + R_S + R_\varepsilon$.

Like Mosier (1939, p. 33) one can say that one of the steps in factor analysis aims at the following:

> In the application of the theory to the factorization of experimentally obtained correlation matrices, the problem becomes one of replacing the experimental matrix, R_1, with unknown diagonal entries, by a reduced matrix, R_0, with known diagonal entries and "true" correlation coefficients.

Thus, in factor analysis, the main attention is concentrated on the matrix, $R = A_m \phi A_m'$. This results in that factor analysis will work with correlation matrices R which, in the principal diagonal, include values corresponding to the common factors of each test.

This means that one relates the correlation matrix, $R_1 = R + R_S + R_\varepsilon$, to m common factors, n specific ones, and n error factors. The specific and error factors, however, can be considered as n unique factors. In the principal diagonal of matrix R one presupposes those values, the so-called communalities, that correspond to the share of the respective test variance belonging to the common factors. The formula for the communality h_j^2 in the jth test is

$$h_j^2 = \sum_{k=1}^{m} a_{jk}^2 + 2\left(a_{j1}\, a_{j2}\, r_{F_1 F_2} + \cdots + a_{j(m-1)}\, a_{jm}\, r_{F_{(m-1)} F_m}\right). \quad (40)$$

The communality is thus equal to the total test variance, with the omission of the unique part of the variance. An alternative way of indicating the communality is therefore

$$h_j^2 = 1 - (a_{jS}^2 + a_{j\varepsilon}^2). \quad (41)$$

This relation has its correspondence in the communality matrix

$$H = I - (R_S + R_\varepsilon). \quad (42)$$

A weakness in many factor analytical methods is that they require knowledge in advance of the matrix, H. There is, as will be shown later, no satisfactory method of estimating H beforehand.

The relation $R = A_m \phi A_m'$ lacks a unique solution even in the case when $\phi = I$. Limiting conditions which give unique and at the same time meaningful solutions must be included. These conditions will be discussed later. In factor analysis it is not always possible to make an exact calculation of the factor values F_{ki}, for the N different individuals because of the fact that the $N(m+n)$ factor values are more numerous than the number of test scores $(N \times n)$.

The unique part of each test variance is included in the matrix

$$R_U = R_S + R_\varepsilon = R_1 - R. \quad (43)$$

If the reliability coefficients, i.e. the values in the matrix, $I - R_\varepsilon$, are known and then consequently also R_ε, the unique variances can be divided into specific variance and in error variance according to the formula

$$R_S = (R_S + R_\varepsilon) - R_\varepsilon. \quad (44)$$

Respective factor matrices can then be calculated according to the formulae

$$A_S = R_S^{\frac{1}{2}} \quad \text{and} \quad A_\varepsilon = R_\varepsilon^{\frac{1}{2}}. \quad (45)$$

The number of factors will be $m + 2n$, if this distinction is made.

In this monograph we shall mainly dwell on factor analysis. As a rule, in psychological applications this method, as distinguished from the component analysis, gives a considerably lesser number of common factors m than the number of tests n. In addition, factor analysis has considerably greater possibilities than component analysis of finding common factors which are invariant in content from investigation to investigation. By suitable selection of tests, as will be discussed in chapter VII, the solution can by factor analysis be found in such a manner that the common factors remain unchanged, e.g. when partly different tests are used or when the population is changed. On the other hand, in component analysis the possibility of getting invariance is considerably less, since the components are hybrids between the common and unique factors found by factor analysis. The content of these components will, therefore, easily be displaced after exchange of certain tests for others which have other unique factors. Neither has anyone seriously maintained that components found by the methods of component analysis now available, are invariant after changes of the test battery's composition (Guilford, 1954, p. 478, and Kendall, 1950, pp. 65–66).

The Geometrical Theory and Rotation

With each one of the various factor methods, one expects to find a factor pattern fitted with characteristics particular to the method concerned. Some of the methods are at present such that it is not possible to find the intended factor pattern by a direct solution of the equation system

$$R = A_m \phi A_m'. \tag{46}$$

Therefore, the analysis must be made step by step. First, a preliminary matrix, A_0, is calculated. It is selected so as to be orthogonal, i.e. follows the relation

$$R = A_0 A_0' \tag{47}$$

and if possible has the number of factors equal to the rank in the matrix R. The matrix A_0 is then transformed to a matrix A_m.

This transformation, which in factor analytical terminology is called rotation, is made with the help of the transformation matrix Λ according to the formula

$$A_m = \Lambda A_0. \tag{48}$$

The transformation matrix Λ should be chosen so that A_m fulfils the relation

$$R = A_m \phi A_m' \tag{49}$$

and, if possible, includes the characteristics that the investigated problem requires. The rotation is often performed graphically by the use of geometrical theory. Geometrical models are, on the whole, valuable aids in understanding and carrying out the factor analytical operations.

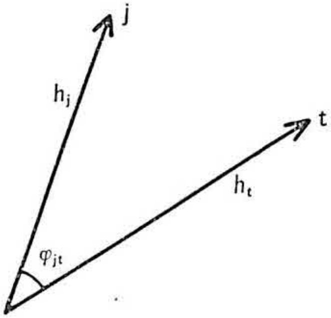

Fig. 1.

The common part of a test j is then in the geometrical model represented by a vector j, the length of which is h_j, the square root of the communality. The test t corresponds to the vector t of the length h_t. ϕ_{jt} is the angle between the vectors in fig. 1. The scalar product of the two vectors j and t is $h_j h_t \cos \phi_{jt}$. The scalar product is equal to r_{jt}, the correlation between the variables j and t. In this way one can demonstrate the correlation matrix R as a cluster of vectors in m dimensions. Factor analysis means that in such a vector system is included a system of reference vectors each with the length of one unit. The test vector system can then be referred to and described in terms of these reference vectors. In the case of orthogonal factors, this is done as follows:

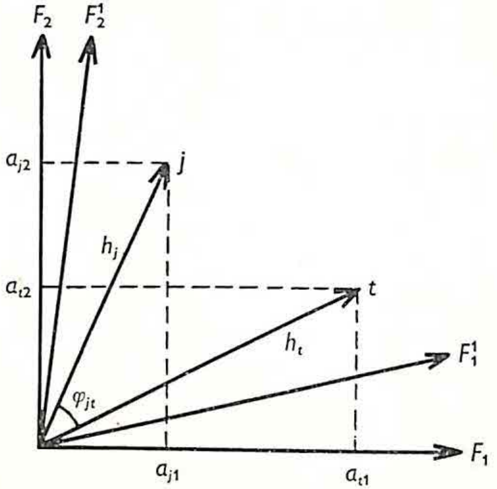

Fig. 2.

The common factor variance in the two tests in fig. 2 is to be referred to the factors F_1 and F_2. Vector j:s coordinates are a_{j1} in factor F_1 and a_{j2} in factor F_2. The coordinates correspond to loadings and will follow the formula

$$r_{jt} = a_{j1} a_{t1} + a_{j2} a_{t2}. \tag{50}$$

The case of correlated factors is more complicated but in principle similar.

The localizing of the reference vectors can be done in many ways. For each correlation matrix there are an unlimited number of possible positions for the reference vectors. In other words, there is no unique solution of a factor analytical problem from a purely mathematical viewpoint, but many different factor patterns are possible. As previously mentioned, certain restrictions must be included in A_m in order to get a unique solution. The choice of these restrictions is one of the major problems in factor analysis.

Another aspect of the above characteristics of the reference vectors is that various solutions, for example bi-factor and centroid solutions, can be transformed into each other through rotation.

This rotation means geometrically that one transforms a preliminary system of reference vectors A_0, established, for example, through centroid solution, so that it changes into another reference system A_m to which have been added certain restrictions. The transformation matrix Λ contains functions of the angles by which the various reference vectors are rotated.

A new reference system can also be selected so that the factors will be correlated—a so-called oblique solution. This shows itself in that the coordinate axes F_1^1 and F_2^1 in oblique rotation, as is seen in figure 2, do not form a right angle.

Assumptions concerning the Nature of the Data

If the above mathematical theory is to be applied to psychological data with meaningful results, certain assumptions must be fulfilled, at least approximately. These assumptions have been discussed in detail, for example, by Thurstone (1947, pp. 55–58) and Wolfle (1940, pp. 2–4). The most important assumptions seem to be the following:

(*a*) Individual differences in a great number of tests are possible to describe as functions of individual differences in a limited number of fundamental variables. In other words, the differences between individuals in a number of tests reflect an underlying structure, so constituted that the behavior tested can neither be referred back to only one undifferentiated trait found in each situation, nor only to a very great number of fully specific traits.

(*b*) The tests have different factor structures i.e. the different factors contribute in various degrees in the different tests.

(*c*) The interaction of the factors in the different tests are possible to describe approximately through addition.

(*d*) The relations between factors reciprocally and between tests and factors are approximately linear.

(*e*) The same pattern of loadings applies to all individuals in the population and the differences between individuals depend on different numerical values in factor scores. The loadings state the average weight given the factor concerned, while the individuals' deviations from the average are given by their factor scores.

The validity of the above assumptions is, at least in principle, testable by empirical methods. If contrary to the assumption *a* the individual differences in the test could be traced back to one undifferentiated trait, then, after correction for attenuation, the correlation between the tests should be equal to 1. If the traits are completely specific the correlations should be very close to zero. In a number of correlational investigations of tests, the correlation matrix has proved to include correlations of varying sizes and structures. Hence, the validity also of assumption *b* that the test can show different factor structures. No scientist, well orientated in factor analysis, doubts the validity of the assumptions *a* and *b* for a number of data in the behavioral sciences.

On the other hand, there is a dispute about the validity of assumption *c*, i.e. that the factors combine by addition. A common objection in psychological applications is that the different sides of the personality act together in a dynamic way that cannot be descibed by addition. Thurstone, for example, answers to this that no one should look for factors which are completely isolated from the rest of the personality for the rest. He points out that if some structure or relation exists, then it must in some way be noticeable and also possible to identify (Thurstone, 1947, p. 59). Besides, further analyses and investigations will have to show whether analysis of factors is possible in a field and if the results are meaningful (op. cit., p. 340). However, one ought to count on the fact that it is not always possible to break down the performance on any test into a sum of simpler functions.

The assumption *d* refers to the linearity between variables. This is also an empirical question that must be decided from case to case. Often one can get linearity by the use of a suitable scaling procedure, e.g. a general transformation to T-scores. If linearity exists between the tests, it is often reasonable to assume at least approximative linearity for the factors. In any case an analysis *can* be made into a number of descriptive factors, which are linearly connected with each other and the tests. Factor analysis should be used in less explored fields where it is necessary to get first approximation and for that purpose linear factor equations can often be used. Further experiments will then be basis for more exact formulations (op. cit., p. 56).

Even the truth of the assumption *e* about the generality of the factors has been strongly questioned. Many psychologists point out that different people can obtain the same score by different approaches, i.e. that their results require wholly or partly different factors. This implies that the same factor equation cannot be applicable to all individuals in a certain population (Allport, 1945, pp. 242–248). As a matter of principle, this only means that the factor pattern becomes more complicated and gives place to more factors than those that should be used if all individuals applied the same procedures for solution of the task. The risk remains, however, that if several procedures for solution, each requiring a very different factor structure, are used, then the loadings in all the different factors will be so small and so many that they cannot be found by the use of such a crude method as factor analysis. Also in this respect the results in each particular case determine whether the obtained factor pattern is applicable and meaningful for all individuals.

From the above discussion, it is evident that the factor equation

$$z_{ji} = \sum_{k=1}^{m} a_{jk} F_{ki} + a_{jS} S_{ji} + a_{j\varepsilon} \varepsilon_{ji} \qquad (51)$$

is often approximately useful as a descriptive model. Further investigations will, in each particular case, determine to what degree this is true.

CHAPTER II

THE MOST IMPORTANT FACTOR MODELS

The Need for Restrictive Assumptions

If the above assumptions are valid for the data investigated, then factor analysis can be used in finding a model that describes the relations between the tests in the form of a few relatively independent factors. There are, as previously mentioned, many alternative factor analytical models, which all reproduce the correlation structure in a certain set of data. The scientist must, therefore, choose which one of all possible models he wants to use. This choice is dependent on his requirements on the factor structure and is based upon more or less explicit theories about the structure.

In the following pages are discussed some of the most important schools of factor analysis and the characteristics of their different factor models. This is not an attempt to give a full historical account but only to bring out aspects of the different schools in factor analysis particularly interesting in this connection.

The Forerunners of Factor Analysis

During the later part of the 19th century there was a debate in Britain, and also on the Continent, about the presence of a general ability and special abilities in the cognitive field. In this debate Spencer and Galton, among others, took part. Spencer maintained a theory of general factors that has great similarity to Spearman's two-factor theory, later proposed while Galton was also inclined to accept what is called group factors in factor analytical terminology. Galton also looked for support from correlation studies in a manner which foreshadowed factor analysis.

During the development of the correlation method that took place at the turn of the century, Edgeworty and Pearson developed the factor equation in its multiple form with orthogonal factors. They worked with anthropometrical measurements and wanted to condense the essential information in a larger number of measurements of different parts of the body to a few independent characteristics. This was done through the principal axes solution. Since they worked with the whole variance in their variables, i.e. with correlation matrices that had ones in the diagonal, the number of factors became the same as the number of variables. The method is thus a form of component analysis, but through this contribution, that came before Spearman's, the formulation of the equations of factor analysis had started. The beginnings of factor analysis has been described, among others, by Burt (1949a).

The Spearman School

It was Spearman (1904) who, through his paper "General intelligence objectively determined and measured", seriously put life into the development of the factor analysis. He introduced his factor analytical method to support one of the theories of current interest about the nature of cognitive processes. Spearman adhered, as previously mentioned, to the monarchic position. In Spearman's formulation it was called the Two-Factor theory, meaning that results in cognitive variables are referred to a general ability factor that in greater or lesser extent is found in all cognitive variables, and in addition, one unique factor in each variable. In his paper, Spearman (1904, p. 284) described his theory as follows:

> All branches of intellectual activity have in common one fundamental function (or group of functions), whereas the remaining or specific elements of the activity seem in every case to be wholly different from that in all the others.

The factor equation for the test j is

$$z_{ji} = a_{j1}F_{1i} + a_{jU}U_{ji}. \tag{52}$$

Spearman does not make a complete distinction between specific and unique factors and often uses the name specific factor for both. In certain cases he works with correlations corrected for attenuation. In this case the specific and unique factors in a test coincide since no error factor is included in the test variance.

The above factor analytical model was included as a part of the so-called Noegenetic Theory about the cognitive processes, which, according to Spearman, was supported by his factor analyses (Spearman, 1931, pp. 112–117). Spearman clearly understood that his correlation matrices could be referred back to other types of factor models, but considered, however, that his model was the most comprehensive one and statistically the most plausible. In addition, the two-factor theory provides simple and unique solutions (Spearman, 1927, pp. VI–VII).

To his death, in 1942, Spearman maintained his conception about the dominating importance of the *G*-factor, which is evident from the posthumous book *Human Ability* published by Jones in 1950. The chapter "Existence of '*G*'" thus concludes with the following statement (p. 61):

> We have marshalled the recent evidence as to whether or not the general factor designated as *G* may be properly said to exist. By this is meant that it presents at least the cardinal statistical characters—definiteness, comprehensiveness and stability—which are indispensable for the purpose of science. The answer, recorded in an immense amount of such research as that of Stephenson, Brown, Alexander and Holzinger, has been a decided affirmative.

Spearman does not enter into any detailed discussion about what he means by the "existence" of a factor. He states that such a question has strong metaphysical elements. In *Human Ability* he and Jones say about the *G*-factor (p. 52):

> Does such a factor really exist? If so, what is its nature? Taking the former question first, let us clear the ground by repudiating all the metaphysical implications that some authors have sought to read into the terms *really* "*exist*". Such topics belong, it is believed by the senior writer, not so much to empirical as rather to philosophical psychology, with which we need not here trouble ourselves. (He has treated it freely elsewhere, as in Psychology Down the Ages.) For us, the question

as to the observation of *G* reduces itself to that of validity of certain "hierarchic" equations. We have only to inquire whether or not these have been sufficiently well satisfied for the purpose of determining and stabilizing the factors at issue.

The existence of the *G*-factor will, according to Spearman, be dependent on that the correlation matrix has the rank one. The two-factor theory is evidently to be considered as a hypothesis that has to be statistically verified by analysis of suitably selected test batteries.

In any case, if the rank is one in the correlation matrix *R*, then there is a general factor, that also is invariant under certain conditions. The latter is an important requirement for a factorial solution, if it is to have scientific value. That a factor is invariant means that its content is unchanged even if the design of the factor study is changed, for example, by exchange of certain tests for others in the battery. A general factor in a matrix of rank one will, therefore, be invariant when adding tests measuring the same general factor and unique factors. This is evident from the following discussion.

If a correlation matrix *R* can be described by a general factor only, then the equation for general factor loadings a_{e1} is

$$a_{e1}^2 = \frac{r_{ej} r_{et}}{r_{jt}} \qquad (e, j, t = 1, 2 \ldots n;\ e \neq j \neq t). \tag{53}$$

A new test $n+1$ has the loading $a_{(n+1)1}$ in the general factor and the correlation $r_{e(n+1)}$ with the test e. The squared loading "a_{e1}^2" determined with the help of the new test will be

$$\text{``}a_{e1}^2\text{''} = \frac{r_{ej}\, r_{e(n+1)}}{r_{j(n+1)}}. \tag{54}$$

When the loadings, under the assumption of one common factor, are included, this will give the equation

$$\text{``}a_{e1}^2\text{''} = \frac{a_{e1}\, a_{j1}\, a_{e1}\, a_{(n+1)1}}{a_{j1}\, a_{(n+1)1}} = a_{e1}^2. \tag{55}$$

By this, the loadings are shown to be unchanged after addition of the new test.

The investigation of the *G*-factor theory will, of course, give an incorrect result, if the tests are so selected that some other general factor fully or partially replaces Spearman's *G*-factor. If, for example, all tests are very time-limited, this can result in a mixture of Spearman's *G*-factor and a speed factor. The factor equation, in such a case, will get the following form

$$z_{ji} = a_{j1} F_{1i} + a_{j2} F_{2i} + a_{jU} U_{ji}. \tag{56}$$

In the analysis, these two general factors can appear as one. Spearman has tried to get away from this weakness by requiring that certain key tests, which experience has shown to be highly loaded with *G*, shall be included in the battery.

Spearman's factor analysis has the characteristic of testing the *G*-factor hypothesis on a test battery measuring the *G*-factor, to which the new tests are added. If the hypothesis is correct, the addition of a new test should not result in a change in the factor structure of the old tests measuring *G* and at the same time the two-factor equation

$$z_{ji} = a_{ji} F_{1i} + a_{jU} U_{ji} \tag{57}$$

will be applicable to the new test.

The ultimate criterion of this is the tetrad criterion by which one can verify that the correlation matrix has rank one. In practice, it was difficult to get the tetrad criterion fulfilled when Spearman and his followers started to expand factor analysis to new and more tests (Burt, 1949 a). But Spearman's defence went mainly along two lines. He maintained that the test battery could include tests that are too much alike and consequently have certain parts of their specific factors in common. These specific factors then manifest themselves in the analysis as group factors, i.e. factors that are included in more than one test but not in all of the tests in the battery. For that reason he stated that the tests that gave rise to such false group factors, so-called "overlapping specifics", should be dropped (Spearman, 1927, pp. 150–160). In opposition to this, the critics, in particular Thomson, maintained that in such a case the scientist starts out from what he wants to get verified and changes the test battery until the results verify the hypothesis (Thomson, 1933).

The second line of defence was of more statistical nature. Spearman considered that the fact that the tetrad criterion was not fulfilled could be caused by sampling errors in the correlations. Therefore, he and Holzinger worked out a formula for the sampling distribution of tetrad differences (Spearman, 1927, p. X). This formula was a decisive progress, but many misused it. They treated individual tetrad differences and investigated whether their size was greater than a certain number—for example, three—times the standard deviation of the sampling distribution. If it was not greater, the tetrad criterion was considered to be fulfilled and the two-factor theory as verified. Particularly, with a small sample, this was easy to attain. This method of proceeding was branded by the critics (Emmett, 1936), and instead, one started to investigate complete distributions of tetrad differences.

As the requirements for test construction and statistical precision increased, it became more difficult to maintain the two-factor theory in its original form. The last major attempt to show its validity was made by Brown and Stephenson (1933). They investigated a sample of 300 10-year-old boys with 22 carefully selected tests. In order to get the tetrad criterion fulfilled, they had to eliminate what they considered to be a verbal factor and two "false" group factors with the use of partial correlation. In addition, they removed one test that distorted the tetrad criterion. The investigation was exposed to much criticism (Thomson, 1933). The debate about this and similar investigations and the development of multiple factorial analysis has resulted in that the presence of group factors in the cognitive field is now generally accepted. Even Spearman had to accept them as facts although he tried to minimize their importance and was inclined to consider them as products of the environment (Spearman, 1927, p. 242). The essential point in Spearman's effort, however, is to determine the presence of and the extent of the *G*-factor, and to find good measures for this. He also considers that he has obtained proof of the central role of the *G*-

factor and that the difference between universal factors and non-universal factors is primary and fundamental (Spearman, 1933, p. 599).

Spearman's co-worker Holzinger has modified the two-factor theory to a so-called bi-factor theory, in which there is also place for a group factor in each test. The factor equation, in this case, is as follows:

$$z_{ji} = a_{j1}F_{1i} + a_{jk}F_{ki} + a_{jU}U_{ji} \quad (k = 2, 3, \ldots m + 1). \quad (58)$$

Holzinger uses his method primarily for testing hypotheses regarding the presence of certain group factors. He builds up his test battery so that it contains three or more tests, that can identify each of the hypothesized factors (Holzinger–Swineford, 1939, p. 1). The factor analysis itself takes the form of a verifying of the hypotheses, but it can also result in that these are rejected and eventually revised for the benefit of another bi-factor solution, which then ought to be tried out on new data.

Holzinger thus tried to refer the correlation matrix R back to a pattern that includes in each test, one general factor, one group factor common to at least two of the other tests, and one for the test unique factor. The group factors can also be fully excluded in certain tests. According to Holzinger, this model satisfies the requirement of simplicity. The factor structure can, for example, look as follows (for the sake of simplicity illustrated only by the sign of the common factor loadings):

TABLE 1.

Test	Factor			
	I	II	III	IV
1	+	+		
2	+	+		
3	+	+		
4	+		+	
5	+		+	
6	+		+	
7	+			+
8	+			+
9	+			+
10	+			+
11	+			

All tests are loaded with the general factor I, and tests 1, 2, and 3 also with the group factor II, and so forth. Test 11 is loaded only in the general factor.

The general factor loadings are calculated by applying Spearman's tetrad principle to all combinations of tests that do not have any common group factor (Holzinger–Harman, 1941, pp. 116–121). Then a calculation is made of the residual matrix after the removal of the contribution of the general factor from the matrix R. If the hypotheses regarding the group factors are correct, the residual correlations between all tests that are not expected to measure the same group factors, should be zero within the limits of sampling errors. The next step is to calculate the factor loadings for the group factors. This is done by applying the tetrad principle to each one of the groups of tests that is supposed to measure a certain group factor.

Holzinger is searching for a certain amount of invariance or "stability" as he often calls it. This invariance is dependent on the tests being so selected that they group themselves into clusters of tests, each cluster corresponding to a group factor expected. The invariance that Holzinger can expect, is thus very limited. If the centroid of a cluster is changed by the addition of new tests, then the content of the corresponding group factor is also changed.

The bi-factor solution, however, is invariant, if new tests are added to the battery, which have the same common factor structure as certain of the earlier tests. This can be shown in an analogous way as in the above case with only one general factor. The invariance is closely connected with the fact that the bi-factor solution is a successive application of the tetrad principle.

An example of the use of the bi-factorial method is given in the study *The Stability of a Bi-factor Solution* published by Holzinger and Swineford in 1939. In this study, the factors found in two samplings from partly different populations correspond well with one another and with the hypotheses. However, Swineford (1949) later tried an alternative bi-factor solution on the same material with success and used this new factor model once more on new material.

Very few parallel investigations have been made with the bifactor method, and therefore it is difficult to appraise the possibilities of the method of giving factors that are invariant after great variations have occurred in the tests and populations investigated (Cattell, 1952, pp. 145–146).

Burt's Factor Analytical Methods

Burt has been in on the factor analytical debate from the beginning. In his major work, *The Factors of the Mind*, published in 1940, he has discussed the theoretical basis of factor analysis even from the viewpoint of general science and philosophy. In his book he treats factor analysis mainly as an aid in finding useful descriptive concepts but will not give these any special meaning in terms of underlying structure (Burt, 1940, p. 227):

> Rigorously speaking, factors cannot be regarded as substances, or as parts of a substance, or even as causal attributes inhering in a substance. They are not separated "organs" or "isolated properties" of the mind; they are not primary "abilities", "unitary traits", "mental powers or energies". They are principles of classification described by selective operators.

He considered the selection of factors as arbitrary in principle, and compared the role of the factors with the network of meridians and parallels of the globe. This network has no meaning beyond the descriptive-mathematical, but it is still very useful for describing the position of any point on the globe. The same is the case with the factor analytical model. It can give psychology as well as other sciences a few permanent, and independent terms of reference.

Burt treats factor analysis as a deductive method, and considers factors found by this method as systematized descriptions, which are the result of certain mathematical manipulations with the data. He emphasizes that this deduction is not sufficient to establish that the obtained factors are fundamental.

In later years Burt seems to be more inclined to use factor analysis for solving real structure problems if used for testing hypotheses.

Burt uses factor analysis as an aid in verifying a "hierarchical theory" as to the nature of the cognitive processes. This theory implies a model containing one general factor and a hierarchy of group factors of decreasing importance. Peel (1953) has given a review of this method and its application. Burt has also supported his theory by data from non-factorial research. In his interpretation of these factors he obviously regards them as more than descriptive concepts (Burt, 1951, p. 47):

> Moreover, this hierarchical structure appears to be something more than a mere logical structure, a mere descriptive device; it would seem to be the outcome of a developmental process, resting on a genetic foundation, and doubtless corresponds to a similar structural organization in the nervous system itself. But a concrete interpretation in causal terms, like all such interpretations, can never be inferred from tables of correlations and matrices of factor-measures alone. To establish it we must introduce non-statistical evidence at some stage of the argument.

Burt, in this case, factor analyzes according to hypotheses advanced beforehand, which have decided the choice of tests (Burt, 1950 a; Moursey, 1952). The mathematical analysis starts with the calculation of a centroid solution. The hypotheses can be such that the centroid solution will get, for example, the expression given in table 2.

TABLE 2. *The centroid solution.*

	Factor			
Test	I	II	III	IV
1	+	+	+	
2	+	+	+	
3	+	+	−	
4	+	+	−	
5	+	−		+
6	+	−		+
7	+	−		−
8	+	−		−

In any case, when investigations are made in the cognitive field, the rotation to positive loadings is made according to the hierarchic division that is indicated by the positive and negative

TABLE 3. *Burt's Group factor solution.*

	Factor						
Test	I	II	III	IV	V	VI	VII
1	+	+		+			
2	+	+		+			
3	+	+			+		
4	+	+			+		
5	+		+			+	
6	+		+			+	
7	+		+				+
8	+		+				+

loadings of the centroid solution. Even the negative loadings are rotated to positive factors, which, in relation to other factors, are orthogonal. This results in an increase in the number of factors. The hierarchical structure in such a case can have the sign pattern indicated in table 3. Burt considers that a hierarchical structure of this type is preferable, because it is simple and corresponds with psychological principles (Burt, 1949b).

In certain cases, for example, when he factorizes personality variables, Burt uses the centroid solution without rotation, since bipolar solution with both positive and negative loadings in his opinion is reasonable in this domain.

Burt's methods have been criticized by Thurstone, Thomson, Cattell, and others, for not giving invariance. Cattell says, for example (1952, p. 249):

> A last principle of rotation is not to rotate at all! Such an economical procedure has its devotees. The most systematic defence of the procedure has been made by Burt, followed by Eysench. The only substantial claim yet put forward, apart from economy, is that it can be empirically shown to give factor invariance just as well as simple structure or any other adequate method. On grounds of experience this cannot at the moment be adequately proved or disproved, since insufficient repeat factorizations of the same variables under adequate experimental and sampling conditions exist today. Those which *do* exist, in the writer's opinion, show more indubitable invariance for simple structure than for nonrotation. However, on theoretical grounds, the use of the unrotated matrix, involving a first general factor, with no zero loadings, does not in the first place agree with the general psychological expectation that

> all factors will be of the same general species of loading pattern and the same general order of mean variance. Further—and this is more surely fatal to the notion that such factors can be invariant—one has to recognize the statistical fact, that, except for one unique situation (when the hierarchy of correlation is not upset), the addition of new variables to the battery alters the centroid of those preëxisting. Their loadings in the first (and in all subsequent) factors vary according to the company they keep. This makes the use of unrotated factors psychologically and statistically meaningless.

This critisism of Burt's methods is somewhat exaggerated, since Burt in certain cases uses rotation. The important thing, however, is that the rotation in these cases is done according to the pattern indicated by the position of the negative and positive loadings of the centroid solution.

The centroid method, which is the foundation for Burt's solutions, is in itself not invariant. It is well known that centroid solutions are changed if new tests are added. This is evident if a test is added to the solution in table 2, which has the same common factor structure as, for example, the first test. Then factors II and III no longer fulfil the requirement in the centroid solution that the sum of the loadings of each factor, except the first one, is equal to zero, since each of the columns II and III has received a positive addition. A new centroid solution should give other loadings. Burt does not seem to count on invariance after such changes in the test battery, as previously mentioned. But on the basis of his earlier studies in the same field, he can select the tests so that a new solution becomes invariant in relation to the solution obtained in the earlier studies. This seems to be an art of balancing the choice of the tests in the battery which requires practice and skill.

Thomson's Contribution

Thomson appeared early as Spearman's foremost critic and has in recent years fulfilled the same function concerning Thurstone. Thomson's contribution to the factor analytical theory discussion has been of a predominantly critical nature, but he has also contributed to the development of the mathematical theory.

In his main factor analytical work *Factorial Analysis of Human Ability*, Thomson, maintains a sceptical attitude concerning the scope of the factorial analysis in the solution of structure problems. He points out that in natural sciences, such terms as "energy", "neutron" and "gene" are considered by scientists as "manners of speech" intended to characterize "abstract ideas" (Thomson, 1948, p. 241). Thomson considers that similar terms in psychology can be motivated by their usefulness, but emphatically points out the danger in giving them any other meaning. He stresses the danger in giving the factors (Thomson, 1948, p. 42):

> ... any more "reality" than an average or an index of the cost of living, or a standard deviation, or a correlation coefficient—though, on the other hand, it may be admitted that some of them, Spearman's *g* for example, may come to have a very real existence in the sense of being both useful and influential in the lives of men.

He has as an alternative model suggested his Sampling theory, which means that a very great number of "bonds" sampled in various ways by the different tests explain the correlations.

The Thurstone School

Thurstone formulated the factor analytical model in a rather general form. He made use of matrix algebra and pointed out that the least number of common factors m that is needed for explaining the correlation matrix R, is equal to the rank r of the matrix.

Thurstone strongly separates two different tasks for factor analysis. He points out in his major work *Multiple Factor Analysis* (Thurstone, 1947, pp. 503–504):

> A factorial analysis can be made for one of two purposes, namely, (1) to condense the test scores by expressing them in terms of a relatively small number of linearly independent factors or (2) to discover the underlying functional unities which operate to produce the test performances and to describe the individual differences eventually in terms of these distinguishable functions. ... It is the second objective that has guided the writer, since 1931, in developing multiple-factor methods. It is the first objective that is usually implied by mathematical statisticians who write on factorial problems. This difference in objectives is largely responsible for controversies about factorial method.

The main object of Thurstone's factor analytical investigation is thus directed on the clarification of the underlying processes. This purpose is, however, linked with a pragmatically-oriented attitude in his theory of science. This theory is further developed in a chapter of *Multiple Factor Analysis* called "On the nature of science", wherein, among other things, he says (op. cit., p. 51):

> It is the faith of all science that an unlimited number of phenomena can be comprehended in terms of a limited number of concepts or ideal constructs. Without this faith no science could ever have any motivation. To deny this faith is to affirm the primary chaos of nature and the consequent futility of scientific effort. The constructs in terms of which natural phenomena are comprehended are man-made inventions. To discover a scientific law is merely to discover that a man-made scheme serves to unify, and thereby to simplify comprehension of a certain class of natural phenomena. A scientific law is not to be thought of as having an independent existence which some scientist is fortunate to stumble upon. A scientific law is not a part of nature. It is only a way of comprehending nature.

The selection of a factor analytical model is to a high degree for Thurstone a question of convenience. It is a question of selecting the simplest and most useful among many models. Thurstone's basis for the selection of model is the principle of parsimony. Starting from this principle Thurstone states a number of requirements to be fulfilled by the factor model if this is to be suitable for expressing the underlying processes.

Thurstone's most important requirements are:

I. PARSIMONY.—According to Thurstone a general rule in scientific theory construction is to use as small a number of concepts with as great a scope as possible. Thurstone therefore tries to find a factor system, which, for example by being simple, may be preferred in the scientific construction of theories.

The parsimony criterion is also used by Thursone for fulfilling another of the requirements he places on the factor model, namely invariance.

II. INVARIANCE.—The factors ought to be invariant in order to be used in scientific models of more general range. This means that factors should not only be dependent on just the one constellation of variables and individuals that are being analyzed at the moment, but also appear in analyses under different experimental circumstances. A verbal factor found in analysis of a test battery in one population, for example, should, in order to be invariant, also appear in a factor analysis of another population with a fully or partially different test battery.

There are different types of invariance which are mutually related. They are dicussed by Thurstone (1947, pp. 360–376), and may be systematized as follows:

(i) *Numerical invariance between analyses of the same battery of tests given to several samples from the same population.* Different independent factor analyses of this kind should give factor loadings which remain the same when allowing for sampling errors. This can be achieved by the use of a factor analytical method giving unique solution.

(ii) *Configurational invariance found in analyses of the same test battery given to several different populations.* In this case the factors should be the same in all analyses, but the numerical size of the loading should be changed according to the differences in the composition of the populations. If, for example, a population has considerably less variance in some of the tests, this will result in a corresponding reduction in the size of the loadings. The numerical values should be changed but the configuration of loadings remain the same. This naturally assumes that the factor structures in the two populations are really comparable. Should many individuals in one population solve the items in the test by means of a procedure which is partially different from that used by the other population, then the results will not be comparable.

(iii) *Numerical invariance between analyses of batteries including partly different tests; each battery given to samples from the same population.* In this case, those tests, which are the same for several batteries, will keep their factor loadings unchanged within the limits of sampling errors. This presupposes, of course, that the variables in both batteries are selected in such a way that the common factors are the same. If, for example, a common factor is loaded in three tests in one battery, and then only one of these tests is included in a second battery, it is to be expected that the common factor becomes specific in this second battery.

(iv) *Configurational invariance between analyses of batteries composed of partially different tests, each battery applied to a different population.* In this case, a test that is included in two partially different batteries will retain its structure, but with changes in the size of the loadings equal to the dissimilarity of the populations. As in case ii, it is here assumed that the common factor variance is similar in both batteries.

If the requirement of invariance will be fulfilled, the variables in the different batteries must be carefully selected. If one battery is built up mainly of very conglomerated variables, e.g. a mixed IQ test or sums of school marks, and the other battery of very homogeneous factor tests, it is difficult to attain invariance between the two batteries.

The invariance problem is further discussed from a mathematical viewpoint in Chapter VII.

III. UNIQUENESS.—In developing his factor analytical methods, Thurstone started from the principle that an infinite number of solutions are applicable to the same factor problem. It is therefore necessary to include certain restrictions to make the solution unique.

For the sake of clarity, Thurstone makes a distinction between two related kinds of uniqueness. The first is of a mathematical nature and is applicable only to one correlation matrix. This kind of uniqueness implies that the limiting requirements should be such, that there is only one matrix A_m which fulfils these requirements and which also satisfies the algebraic relation

$$R = A_m \phi A_m'. \tag{59}$$

The second kind of uniqueness is present if the factors are unique in the sense that they represent the only acceptable model of the underlying processes studied. As previously mentioned, Thurstone adopts a fundamental relativism in regard to this question, and believes that there may be several alternative models which are equivalent. But, if uniqueness in the above sense is desired certain limiting requirements have to be imposed, which result in that only one configuration of factors will occur in repeated investigations with different test batteries within the same domain (Thurstone, 1947, p. 333). This requirement for uniqueness is, as can be seen, closely related to the need for invariance.

The two requirements for uniqueness signify a need of requirements that at the same time can give an acceptable factor model, from a scientific viewpoint, and a mathematically unique solution of the relation $R = A_m \phi A_m'$.

IV. FUNCTIONAL UNITY.—The concept of functional unity is based upon the theory that response patterns to certain types of stimuli are more or less closely related to each other or overlap in the sense that partly the same processes are involved. This grouping of response patterns or their elements may, for example, be the result of the same genetic basis or that the environment has had a similar effect on them. Functional unity can, for example, be characterized in the following way (Cattell, 1946, p. 71):

> ... the unity of a set of parts is established by their moving—i.e., appearing, changing, disappearing—together, by their exercising an effect together, and by an influence on one being an influence on all.

It may, for example, be assumed that results obtained in different kinds of tests for speed of reaction are dependent on a certain center in the central nervous system, which gives rise to a high correlation between the results. A numerical factor, for example, may partly be the result of training in a number of processes at school, which are related and highly dependent on each other. Such interrelated response patterns will operate as relatively fixed groups of processes, called by Thurstone, "functional groupings" or "functional unities".

To what extent factors of the "functional groupings" type are present, and their degree of stability, is an empirical question.

V. AGREEMENT WITH NON-FACTORIAL FINDINGS.—Though it is important that factors show numerical or configurational invariance between different factor studies, the decisive criterion of the value of the factors is, however, the extent to which they are fruitful in non-factorial psychological research.

According to Thurstone, what is needed is a combination of factorial and non-factorial research in the solution of a structure problem. In his opinion, the main use of factor analysis is to give a first outline of a new domain (op. cit., p. 56):

> The exploratory nature of factor analysis is often not understood. Factor analysis has its principal usefulness at the border line of science. It is naturally superseded by rational formulations in terms of the science involved. Factor analysis is useful, especially in those domains where basic and fruitful concepts are essentially lacking and where crucial experiments have been difficult to conceive. The new methods have a humble role. They enable us to make only the crudest first map of a new domain, but if we have scientific intuition and sufficient ingenuity, the rough factorial map of a new domain will enable us to proceed beyond the exploratory factorial stage to the more direct form of psychological experimentation in the laboratory.

Factor analysis and other methods can be used in combination so that factor analysis generates hypotheses that are verified by other methods or so that factor analysis is testing hypotheses formulated on basis of other methods.

In order to find factors that meet the above requirements, Thurstone uses the principle of simple structure. The methods based on this principle are used by Thurstone not only for testing hypotheses but also for generating hypotheses. This procedure is discussed in Chapter VI.

CHAPTER III

DIFFERENT APPLICATIONS OF FACTOR ANALYSIS

From the above historical outline it is evident that opinions are divided regarding the assistance factor analysis can give in solving problems in the behavioral sciences. Here follows an attempt at presenting a more systematic discussion of the types of questions to which factor analysis can be applied and of the methodical problems involved. The major areas are (1) to find the number of dimensions; (2) to find descriptive factors; and (3) to find explanatory factors.

The Problem of Finding the Number of Dimensions

Factor analysis has made important contributions to the behavioral sciences both directly and indirectly by emphasizing and defining more strictly the concept of dimensionality. It has become more evident in many areas that a large number of the variables used are multi-dimensional. The concept of dimensionality is not new of course. It is involved, for example, in some of the theories, which were advanced before this century, as to colour vision and olfaction. But the models and mathematical methods of factor analysis can make these theories more exact and help also in testing and refining them. New fields have been established to a large extent because of influence of factor analysis. Multi-dimensional scaling and latent structure analysis are examples of this.

Here dimensionality problems in the theory of mental test will mainly be discussed. For example, in selection and construction of a test battery for prediction, one wants to find the smallest number of dimensions measured by the tests. Expressed mathematically, this is the same as finding the rank of the matrix comprising the correlations between the tests.

In calculating the rank, two problems may be differentiated, depending on what type of correlation matrix is used:

(1) To find the rank of the correlation matrix R_1 with unities (or possibly reliabilities) in the diagonal. This case is connected with component analysis.

(2) To find the rank of the correlation matrix R with communalities in the diagonal. This case is connected with factor analysis.

In the first case it is a question of finding the smallest number of factors, which can explain the total variances in all tests. When the battery contains mental tests, the number of dimensions is nearly always equal to the number of tests, since each test partly measures something which is specific for it.

In certain cases, however, the correlation structure may be such, that a considerably smaller number of factors than the number of tests almost completely explains the variances and covariances. Such a case is considered, for example, by Guttman (1950) in his scale analysis. By using scale analysis Guttman endeavours to test the homogeneity of a scale. The requirement for homogeneity, according to Guttman, is, that all items in the scale are completely correlated, i.e. that the rank is one of the matrix R_1 with correlations between items and with unities in the diagonal. This implies that all correlations are of size one. This requirement is never fulfilled in practice. Guttman measures homogeneity by the coefficient of a reproducibility, which has its maximum value, unity, when the scale is completely homogeneous. In most cases the degree of homogeneity is considered to be sufficient when the coefficient exceeds 0.9. Certain assumptions must, however, be fulfilled, e.g. concerning the frequences of different answers to the items in the scale.

The testing of homogeneity by Guttman's method is a kind of component analysis. It is examined if *one* component can reproduce practically the total variance in all the different items of the scale. In a homogeneous scale, according to Guttman, all items must measure the same dimensions and differ in the frequencies of various answers. Guttman's discussion refers mainly to the field of attitudes. Similar ideas for the field of power tests have been developed by Loevinger (1947).

In the second case, a correlation matrix R with communalities in the diagonal is considered. The intention here, is to find the smallest number of factors that can describe the relations between the different tests, i.e. the common parts of the variances in the tests. A concept of homogeneity, different from that accepted by Guttman and Loevinger, is related to this case. A scale composed of items, which correlate one with each other, is very difficult to get, especially if the scale should have predictive capacity. Another type of scale, more useful for prediction, is a scale, composed of items which all measure the same general factor, and where each item measures one specific factor. In such a case, matrix R, with correlations between items and with communalities in the diagonal, will have rank one. This concept of homogeneity is mentioned, for example, by Cureton (1951, p. 646).

Even in a scale which is composed of such items and which measures also specific factors, the general factor will have a tendency to dominate the total variance because the general factor of each item has correlation one with the same factor in the other items. The different specific factors are, on the other hand, zero-correlated reciprocally and with the general factor.

For one individual in a test scale composed of n items, each of which is loaded with the general factor and one specific factor, the total score $\sum_{j=1}^{n} z_{ji}$ is expressed in the following factor equation:

$$\sum_{j=1}^{n} z_{ji} = \sum_{j=1}^{n} a_{j1} F_{1i} + \sum_{j=1}^{n} a_{jU} U_{ji}. \tag{60}$$

The total variance of the test scale is

$$s^2 = \frac{\sum_{i=1}^{N} \left(\sum_{j=1}^{n} z_{ji} \right)^2}{N} = \frac{\sum_{i=1}^{N} \left(\sum_{j=1}^{n} a_{j1} F_{1i} + \sum_{j=1}^{n} a_{jU} U_{ji} \right)^2}{N}$$

$$= \sum_{j=1}^{n} (a_{j1} + a_{jU})^2. \tag{61}$$

Since all unique factors are uncorrelated both mutually and in respect of the common factor, the equation is

$$s^2 = \sum_{j=1}^{n} a_{j1}^2 + \sum_{j=1}^{n} \sum_{t=1}^{n} a_{j1}\, a_{t1} + \sum_{j=1}^{n} a_{jU}^2 \quad (j, t = 1, 2 \ldots n;\ j \neq t). \quad (62)$$

It is evident that the general factor will contribute to the size of the test variance also through the term

$$\sum_{j=1}^{n} \sum_{t=1}^{n} a_{j1}\, a_{t1}.$$

For example, when $a_{j1} = a_{jU}$ for all items, the general factor will contribute $n/(n+1)$ of the whole test variance, in spite of the fact that it only corresponds to half of the variance in each item. In such a scale of 24 items, for example, 96 per cent of the variance will refer to the general factor. Analogous relations can be shown to exist when the general factor has other loading values or when several common factors are measured by the test scale. This condition does not seem to have been formalized in this way before.

The rank of the matrix can, of course, be calculated by methods other than factor analysis. The latter procedure, however, has often the advantage that the factor loadings found can then be used in a subsequent analysis of the data, in order to find which the dimensions are.

In testing the rank of a correlation matrix, it is important to determine the confidence limits for the residual correlations, after a certain number of factors are taken out from the correlation matrix. This problem is treated in Chapter VIII.

The Problem of Finding Descriptive Factors

Factor analysis makes it possible to refer the common parts of the variances, in a number of observed variables, to a few factors whose intercorrelations are either low or zero. The common factor loadings reproducing the information in the $n(n-1)/2$ intercorrelations of correlation matrix R, are then concentrated in the $n \times m$ loadings of matrix A_m. There is an unlimited number of matrices of rank m, which are all alternative and intertransformable descriptions of R. Consequently, it is a question of selecting a solution with the most suitable characteristics.

More or less definite hypotheses about the nature of the descriptive model are often constructed. Factor analysis can then be used for testing whether a matrix A_m, which corresponds to these hypotheses, can be fitted to the correlation matrix R. If this is the case it proves that the hypothetical model is an acceptable description. Other considerations and investigations —whether factor analytical or not—will then determine if this or some other factor matrix is the best description for a certain purpose.

In many cases there is no particularly specified hypothesis available that implicates a certain structure of the matrix A_m. The structure has to be selected according to more general scientific principles. These may, for example, be obtained from the special field of research, from statistical theory, or from general scientific principles. A principle often used is to reduce the number of factors m in the matrix A_m to the least possible number, that is the rank r in the correlation matrix, R. An additional method for getting a simple description is to use orthogonal factors, i.e. a solution A_m chosen in such a way that the matrix ϕ, including the correlations between the factors, is a unit matrix.

Sometimes it may be better to maximize, by a principal component solution, that part of the common factor variance of the battery, which is taken out with the first factor, and then apply the same for the second factor, and so on. In this case, it may happen that a number of factors, fewer than rank r, reproduce such a great part of the correlation matrix that the residuals can be neglected, when this smaller number of factors has been taken out.

A principle common to several methods, e.g. the simple structure method and the bi-factor method, is that they shall fulfil the requirement for a certain number of zero-loadings in the matrix A_m. Each of these methods has, as shown in Chapter II, special requirements in regard to the distribution of the zero-loadings.

The above-mentioned and other principles are employed in different combinations in the various factor analytical methods. In certain cases, different principles conflict with each other, and it is necessary to decide which one should be given priority. In the method of simple structure, for example, the requirement for a certain distribution of zero-loadings is given priority over the orthogonality, while for the bi-factor solution, it is rather the opposite.

It is not the intention here to go into a detailed discussion or to give a systematic account of the various alternative methods. This is done, for example, by Holzinger–Harman (1941) and Cattell (1952). It is sufficient to point out that it is a question of a conventional choice, where criteria such as the special aim of the description, its simplicity, and statistical efficiency, are the deciding principles.

Descriptive Factor Analysis and Its Applications

In everyday life and partly also in scientific psychology, a very extensive vocabulary referring to a lot of abilities, personality traits, symptoms, etc. is used to describe human behaviour. An example of this is the list of about 18,000 English terms designating behavior compiled by Allport and Odbert (1936). Many of these words are synonyms but others are intended to denote some shade of meaning associated with a particular word. A description of a human being in a number of categories is, of course, too comprehensive. The list must be reduced to a few essential variables. In other words, it is a question of selecting, from the universe of all possible variables, a suitable group which makes possible a rather complete but still simple description.

Factor analysis is a useful tool for such a reduction, as it gives a description of a large number of variables, in terms of a few relatively independent factors.

Factor analytical research programs, of the type outlined above, have been carried out, for example, by Burt, Holzinger, and Cattell. The factors found in these studies may be regarded as being relatively fundamental, descriptive variables. Whether they can be someting more, is a question to be treated later in this monograph.

Factor analysis can be carried out either as an exploration or as a test of hypothesis. In exploratory studies, the factor analysis is carried out principally without definite hypotheses as to the occurrence of a certain factor structure. The selection of a particular factor analytical method involves, however, a choice of the general type of the factor analytical model. In a solution based on, for example, a general factor, a large part of the total variance in the battery is compressed into a general factor, while in a solution applying the principles of simple structure, a general factor of the first order is avoided.

In hypothesis-testing, factor analysis is made according to theories established in advance that certain tests are loaded with certain definite factors. Factor analysis is carried out in such a way, that these factors, if present, will be apparent. For example, if one wants to test the hypothesis that results in geometry can be described in terms of a general ability factor, in addition to a spatial factor, then a test battery, including ability tests of different types, also spatial, and various tests of achievement in geometry, are selected. The battery is administered to a sample, where individual differences both in general ability and in geometrical aptitude are expected to be present. The resulting correlations are factor analyzed, e.g. by Holzinger's bi-factor method, according to the hypothesis that a general factor is present which is loaded also in the geometry tests. If the hypothesis is correct, a general factor will be obtained which is loaded in all tests, plus a group factor having loadings in the spatial and geometrical tests, but not in other tests, for example, the verbal tests. If this factor model is applicable to the data it shows that in such a population geometrical results can be partly described in terms of a general factor and a spatial factor. On the other hand, it does not signify that this model is either the most suitable, or the only possible one, since there are other factor analytical models that give alternative descriptions. Other criteria, for example the usefulness of the model in test construction and in educational connections, must be used to decide whether the model is the most suitable description of the domain.

The main advantage of factor analysis, as an aid in finding descriptive concepts, lies in the fact that it can help in finding factors that are easier to comprehend than the multiplicity of tests. Instead of carrying out more detailed observations and experiments on a single test, in an attempt to study what it measures, the correlations between a large number of tests may be used for this purpose. The intercorrelation matrix indicates to what extent the different tests measure the same thing, and this in itself is an important basis for drawing conclusions about the nature of the characteristics or factors of which the tests are a measure. Factor analysis of the intercorrelation matrix gives still greater possibilities of comprehending and investigating the nature of the factors measured by the tests, since the factor matrix summarizes, in a few factors, the information included in the intercorrelation matrix. From the factor matrix, it is evident in which test each factor has high and low loadings respectively, and in which tests loadings are lacking. This gives a broad basis for a psychological analysis whose purpose is to identify factors that describe, in a simple way, the connection between the tests.

Finally, it may be stated that through factor analysis one can get suggestions for descriptive models that include a few relatively independent, and objectively established, parameters. Factor analysis can also be used for the strict testing of hypotheses about suitable descriptive models.

Some examples of the application of factor analysis for descriptive purposes will now be given.

A promising field of application for factor analysis is studies of changes in the content of the variance of a certain test, when the individuals are subject to different types of influences, such as growth and practice. The same technique applies when a test is given under different testing conditions, for example, after changes in time limits.

Investigations of this kind may be made in the following way. The test is given to a sample of individuals under different conditions, for example during different phases of a learning process. The results obtained on the different occasions form one variable for each occasion. A battery of carefully selected reference tests measuring well known factors is given to the same sample.

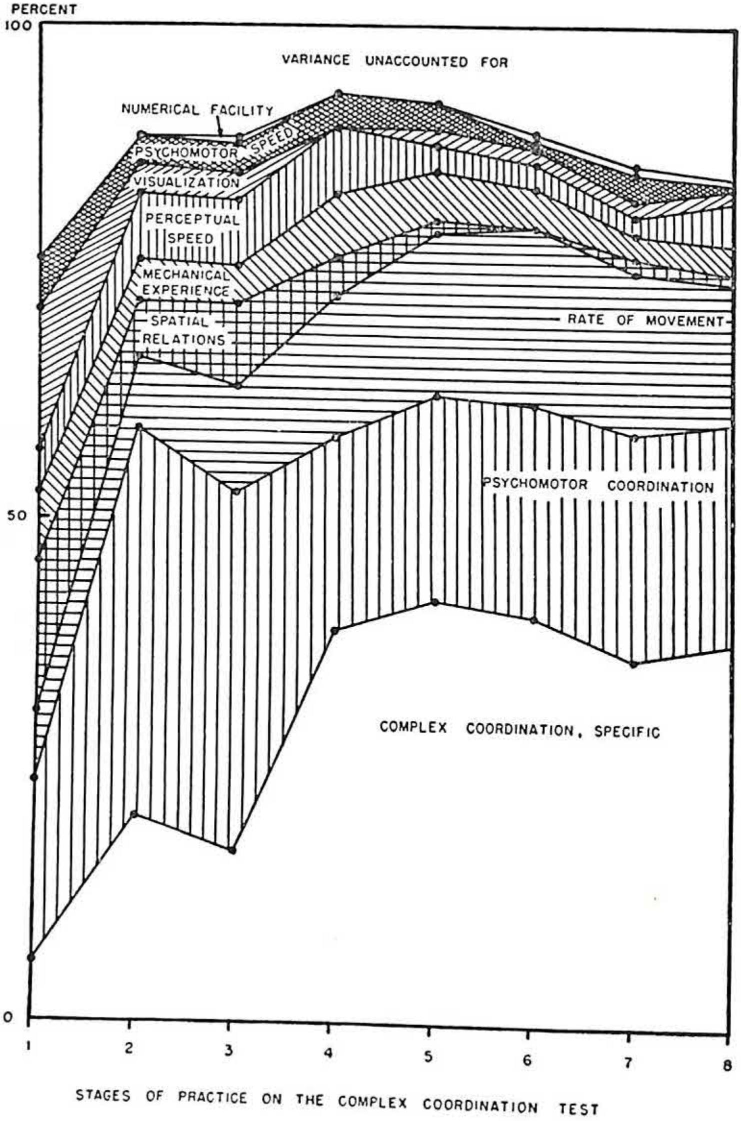

Fig. 3. Percentage of variance (shaded area) represented by each factor at different stages of practice on the Complex Coordination Test. (From Fleishman & Hempel, 1954.)

These factors are expected to be loaded in the variables formed by the test studied under different conditions. Correlations between all the variables are calculated and subjected to factor analysis. The factor solution is rotated according to the factors which the reference tests are expected to measure. By this operation, the variances of the variables, based upon the different testing occasions, will be described in terms of the known factors which are measured by the reference tests and by other factors common mainly to the variables based on different testing occasions. In a factor matrix containing such factors it is possible to get a clear and useful picture of the changes in factor structure from one occasion to the other.

A good example of this approach is Fleishman and Hempel's (1954) study of the effect of training on performance in a complex psychomotor test. It was found that the longer the training continued, the less extensive became the proportion of the variance in the complex psychomotor test, which corresponds to factors measured by the reference tests. This is shown in Fig. 3. The special "psychomotor" factors became more dominant, the longer the training continued. Results of this type have implication both for learning theory and for prediction of abilities that are to be trained.

Another field of application is item-analysis, in connection with the construction of scales, questionnaires, etc. Earlier it has been discussed how factor analysis may be used in the construction of homogeneous scales. The method can also be used in studies of which dimensions are involved, when the matrix R of correlations between items has a high rank. In such factor analyses many special problems occur that are not yet completely solved. They relate, for example, to methods for calculating correlations when the levels of difficulty vary between the items (Vernon, 1950 b and Gourlay, 1951).

Prediction and Factor Analysis

Prediction and the requirements implied by this in regard to the variables, are important psychometrical problems that have been rather thoroughly dealt with from a theoretical viewpoint. A systematic discussion of this topic is found, for example, in *The Prediction of Personal Adjustment* (Horst *et al.*, 1941), where the role of factor analysis is mentioned, and also in *Theory of Mental Tests* (Gulliksen, 1950).

The selection of tests for prediction gives illustrative examples of the problems that arise when looking for suitable descriptive variables, and of the aid factor analysis can render. There are two types of problems of prediction to be discussed in this connection: direct prediction and differential prediction. The purpose of direct prediction is to predict, with the greatest possible validity, the subject's results in a certain field. The selection of personnel for a certain job, or of pupils for a certain type of training are examples of this. Often the same set of predictors should be used for direct prediction in many fields at the same time; this is the problem which will be discussed here. For this purpose, a number of tests[1] are needed, which are chosen in such a way that they give, in a simple manner the best possible predictions of results in all the jobs mentioned. From a psychometric viewpoint, the problem may be said to consist in finding a limited number of tests that, weighted together with suitable weights, give maximum correlations with criterions of success in respective jobs. The problem of differential prediction arises when a limited number of people are to be given different jobs according to assigned quotas, so that the total sum of job results is maximized. This problem will be considered later.

The methods for planning, construction and selection of a test battery to be used for direct prediction of success in a number of jobs simultaneously, are relatively well developed. Therefore, it is sufficient, to mention the different steps briefly and concentrate the discussion on the contributions which can be made by factor analysis.

(1) JOB CLASSIFICATION.—The aim of this procedure is to find the jobs that are similar in important and measureable

---

[1] The argument is, in principle, applicable even to other types of observed variables, for example biographic and physical data, but is here applied to tests only. Further, the assumption is made that all tests are equal from administrative points of view such as cost and testing time.

respects and can be grouped together in so-called job families, in which approximately the same traits are required for success. Factor analysis can assist in making such a grouping simpler and more objective. Coombs & Satter (1949) have made such a factor analysis of job descriptions using correlations based on the number of common elements in the descriptions of each job. The dilemma in grouping jobs in job families is, that if many jobs are assigned to one family, what is typical of the different jobs will, to a certain degree, be neglected in favour of more general properties. If, on the contrary, one goes to the opposite extreme and keeps the special jobs almost completely separate, the number of job families will become too large to handle conveniently and at the same time many of the properties will be very specific. Hence some kind of compromise between the two extremes must be made. Grouping in job families cannot be kept separate from the job analysis, which is the next step.

(2) JOB ANALYSIS.—This step aims at analyzing job results into, if possible, measurable categories, and at determining how important each one of them is for success in the job. A number of different methods may be employed for job analysis (Kerschner, 1955). Factor analysis has also been used in this connection, primarily for systemizing and limiting the number of factors in the job descriptions. McQuitty *et al.* (1954) have, for example, factor analyzed about 150 descriptive phrases that have proved to differentiate between good and poor flight mechanics, and have reduced these phrases to 23 factors.

(3) TEST CONSTRUCTION.—The success of this is, of course, to a great extent, dependent on the quality of the job analysis. The selection and construction of the tests can be made primarily along two lines, namely, in the form of analogous tests or of analytical tests or of some compromise between these two types. Analogous tests are constructed, in such a way as to present the testees with tasks that are similar in a limited form, to the tasks that are later to be carried out in the job, i.e., they are analogous to the job requirements. An example of this is a radio-code-learning test in which the subjects, after having learned some simple Morse-letters, are tested in their ability to receive these

letters. An analogous test is often an effective predictor for exactly the job it is intended for, but generally has a poor predictive capacity for other jobs because of its specific character. In the construction of a test battery for the prediction of a whole series of different jobs, it will often be necessary to have at least one analogous test for each job.

The analytical tests are intended to measure the important elements of the jobs in question. From a practical viewpoint, each of the tests ought to measure an element which is so fundamental that it can be used for several jobs. Tests measuring such fundamental elements will have a wider range and will, on the whole, be more applicable and flexible than analogous tests. The main principle for construction of analytical tests must be to find, by building on the job analysis, elements that have validity for many job families, and then to construct tests for the most important of these elements. By employing analytical tests it is possible to make predictions for a number of jobs, by means of a few tests. For this to be successful, it is necessary for the different tests to have both high validity for many job criteria, and low intercorrelations. The low correlations between the tests show that they measure different elements, i.e., cover different parts of the variances in the job results.

The importance of this is clearly evident if one studies which correlations a new test, called $n + 1$, should have, if it is added to a previously used battery of n tests, in order to increase its predictive power.

The calculation is based on the following correlation matrix:

$$R_1 = \begin{bmatrix} 1 & r_{0(12\ldots n)} & r_{0(n+1)} \\ r_{0(12\ldots n)} & 1 & r_{(12\ldots n)(n+1)} \\ r_{0(n+1)} & r_{(12\ldots n)(n+1)} & 1 \end{bmatrix} \tag{63}$$

In this case $r_{0(12\cdots n)}$ is the multiple correlation between the criterion and the original tests, $r_{0(n+1)}$ is the correlation between the criterion and the new test, and $r_{(12\cdots n)(n+1)}$ is the correlation between the original battery and the new test. The original battery is weighted together with the same weights as against the criterion. The squared multiple correlation between the criterion and the test battery including the new test is as follows:

$$r^2_{0(12\ldots n+1)} = \frac{r^2_{0(12\ldots n)} + r^2_{0(n+1)} - 2r_{0(12\ldots n)}r_{0(n+1)}r_{(12\ldots n)(n+1)}}{1 - r^2_{(12\ldots n)(n+1)}}. \quad (64)$$

By means of this formula the following conditions are found for the increase of the multiple correlation:

If $r_{0(n+1)} > r_{0(12\cdots n)}$ then the multiple correlation will always increase.

If $r_{0(n+1)} = r_{0(12\cdots n)}$ then the multiple correlation increases when $r_{(12\cdots n)(n+1)} < 1$.

If $r_{0(n+1)} = kr_{0(12\cdots n)}$, then $r_{(12\cdots n)(n+1)}$ must be less than k if the multiple correlation is to increase ($k<1$).

Generally speaking, if a new test is able to increase the multiple correlation, then the quotient of the correlation between the criterion and the new test, and the correlation between the original battery and the new test, must be greater than the correlation between the criterion and the original battery, i.e.

$$\frac{r_{0(n+1)}}{r_{(12\ldots n)(n+1)}} > r_{0(12\ldots n)}. \quad (65)$$

A new test can improve prediction through its high validity or through its low correlation with other tests or through a combination of both.

Factor analysis is a useful tool for the development of tests with the above-mentioned characteristics. The construction of a test battery for the purpose of prediction, consists, according to factor analytical terminology, in finding the factor structure of the job results and in designing tests that are as pure measures of these factors as possible.

The role of job analysis as a basis for test construction, and how factor analysis can facilitate such job analyses, has previously been discussed. In order to make job analysis useful, its results must be presented in a terminology that fits the frame of reference in which test results are described. The discrepancies in the terminologies of job analysis and of test construction are unfortunately great. This difficulty can be avoided by factor analyzing criteria of success in respective jobs, together with a test battery composed of reference tests with a factor structure well established through earlier factor analyses. Then the criterion variance will, to the extent that the tests are correlated with the criterion, be described in terms of a number of relatively well-defined and well-known factors that have low intercorrelations. In this case, a whole test battery makes it possible to obtain a better survey of the criterion structure, and a great many hypotheses as to factors of importance for success in the job, can be tested simultaneously. On the basis of the first results, new hypotheses as to the factor structure of the criterion can be advanced; these hypotheses lead to the addition of new or reconstructed tests in the next analysis, and so on, until as much as possible of the criterion variance is covered by well-known factors. Such methods have been used, for example, in the U.S. Air Force; and have been further discussed by Guilford (1948).

The selection of tests to be included in an analysis of the above type is now simplified through the monograph *The Description of Aptitude and Achievement Tests in Terms of Rotated Factors* by French, (1951). It contains descriptions of well established factors and of tests that measure them as well as possible. The purpose of that monograph and a rather optimistic opinion on the capacity of factor tests in prediction, are stated in its introduction (p. v):

> This book is devoted to the progress of test development toward the situation where the test constructor has a file of tests to measure each factor of the mind. Within each file will be hard tests, easy tests, long tests, and short tests, but they will all measure in pure form the factor named on the label of the file. When the test constructor of the future gets an order for a tailormade test, he will consult his summaries of past researches and determine what factors need to be measured and what weights should be used with them in arriving at a test battery having maximum efficiency. Should the problem posed by his client be of so novel a nature that no past researches apply, it will be possible simply

to turn to a file labelled "Ready-to-use Diagnostic Batteries" and select a suitable collection of standard pure-factor tests which reveal the factorial composition of a suitable criterion population.

On the basis of his studies of earlier analyses and his own investigations French has also published a series of tests that measure the various factors as well as possible (French, 1954).

Often there are no criteria available in due time or they are so defective that the preliminary stages of the selection of the test battery must be made without criteria. The selection of tests must then be based primarily on "construct validation" (Cronbach&Meehl, 1955). In this procedure, statistical comparisons with criteria of success are replaced by psychological considerations and indirect proofs, which indicate that the tests measure what they are intended to measure. Factor analysis facilitates such determination of the test structures, especially if the proposed tests are included in a battery of reference tests with a well-known structure. One can then get, in the factor matrix, safer and more objective grounds for the identification of what the tests measure.

(4) THE PROBLEM OF WEIGHTS.—The last stage in the development of a test battery for prediction is, the determination of the weights the test should have in prediction for the various jobs. A method which is theoretically correct, is to determine the multiple correlation between the job criterion in question, and the test battery, and to take the regression coefficients obtained as weights. Factor analysis is not a substitute for this procedure, since factor scores are only weighted sums of test scores, and cannot increase the correlation that is already maximized through the weights obtained in the calculation of multiple correlation (Thomson, 1948, pp. 114–116). In order to work well, the multiple correlation procedure requires, however, an unbiased and relevant criterion. If this is not the case, it becomes more important to base the weights on knowledge of the job requirements and according to the nature of the tests that can be obtained from job analyses, factor analyses, etc.

The determination of regression coefficients by the calculation of multiple correlations is very sensitive for sampling errors.

This has resulted in an emphasis on the requirement for the cross validation of the tests on new samples (Mosier, 1951 a). The fluctuations in the regression coefficients and the drop in the value of the multiple correlation after cross-validation become greater as more tests are included in the battery, and as the samples get smaller. These difficulties have been discussed in detail, for example, by Horst (1941), who points out the importance of reducing beforehand the number of tests that are to be included in the regression equation. He also mentions that factor analysis is an effective aid in this respect. On the basis of factor analysis, a smaller number of pure-factor tests may be substituted for a larger test battery of complex tests.

Efforts have also been made to use factor analysis as a step in regression analysis itself. The aspects of factor analysis, which are found attractive, are the possibilities of condensing the common variances in the correlation matrix R to a matrix A_m including only a few orthogonal factors. Especially when R is to be used in many mathematical calculations, for example, in the estimation of multiple correlations and regression weights, it might prove advantageous to adopt what is often the harder method, and first estimate A_m (Dwyer, 1940, and Guttman, 1940).

One use of the common factor matrix, when calculating regression equations, is for finding a system of weights which in cross-validation gives rise to less shrinkage of the multiple correlation. Leiman (1952) has tried to use the common factor matrix in the calculation of regression weights; and this is his reason for doing so:

> The rationale for this is that, since the number of factors is ordinarily less than the number of independent variables in the system, fewer degrees of freedom will be lost, and if N remains the same, the standard errors of the regression weights will be smaller.

Leiman tested this theory on a population of 1,468. In the first study he divided the subjects into samples of 30, and in the second study into groups of 200. The result was a significantly higher cross-validity for the common factor regression weights in the first case, but they were not significantly higher in the second case, when compared with regression weights calculated in the conventional manner.

A drawback, if the identical results of calculations are required which would be obtained by conventional regression analysis, is that, in practice, the factor matrix A_m never fully exhausts R. There is always a certain residual variance left out, even if it is kept within the limits of sampling errors by the use of a significance test. Another reason against the use of factor analysis in this connection is the rapid development of electronic computers. This will mean that labour and time elements in numerical calculation is less important.

As has already been mentioned, in regression analysis a reduction in the number of tests is required, so as to obtain more stable regression weights after cross-validation. Various methods have been discussed for excluding such tests that either do not increase at all, or increase the size of the multiple correlation only very slightly, as compared with the size it acquires by means of a few of the tests. Dwyer (1940), among others, took up this problem and mentioned that factor analysis can assist in reducing the number of tests. This idea was still further developed by Lubin & Summerfield (1951). They used the diagonal method of factoring for finding those tests, which taken together, are the best predictors. They take as their starting point the test that has the highest correlation with the criterion, and then find a second test which, when added to the first, will, to the greatest extent, increase the predictive ability, and so on. Finally the tests will be left which can only increase the predictive ability of the battery to a very small extent and which may be excluded.

A rougher method, which is often better, however, if the criterion is poor or complex, is to weight together as to "composites", those tests which measure the same common factors. These composites may be relatively few and low-correlated and can be more easily weighted together in a regression equation to stable total scores. This grouping of tests into composites can often be done by inspecting the correlation matrix; but, particularly if the correlation structure is complex, the factor matrix can produce a simpler and clearer picture of the relations between the tests.

Another possibility suggested by Dwyer is to factor analyze the battery, including the criterion, and then to extract one test for each factor that has loadings in the criterion. This, however, presupposes that it is possible to find relatively pure measures of each factor. If this is the case, the multiple correlation between the criterion and the *m* number of factor tests should be rather close to the multiple correlation given by the whole battery of *n* tests, and in addition should possess the advantage of being considerably more stable in cross-validation (Dwyer, 1939).

Both the method of using composite scores and that of single factor tests require that the tests are rather pure measures of the factors included in the criterion. In practice, it is often very difficult to fulfil this requirement adequately, in any case in regard to paper and pencil tests.

Differential Prediction and Factor Analysis

In the previous section are considered the requirements which a test battery has to satisfy in order to afford direct prediction in respect to each of a number of jobs. The object is to obtain the highest possible multiple correlation between a battery of few tests and each of the criteria of job success. In most cases the various job criteria are more or less intercorrelated. Consequently even if relatively uncorrelated factor tests are employed, these will, with weighting systems, which are largely similar, be present in regression equations for the prediction of several jobs or job families. Another result of the high correlations between the criteria, is the well-known fact, that general measures of ability, for example, a general intelligence test, often give rather good prediction for a number of different jobs. If such a test dominates the selection for all these jobs, then practically the same individuals will be considered the most suitable for all the jobs. This is not a drawback from the viewpoint of a pure personnel selection procedure, when the number of applicants is large—in principle, unlimited—in relation to the demand. On the contrary, selection by such generally useful tests can often save time and effort, since prediction to all the different jobs can be made by only using a few tests.

The problem is different if the number of applicants is limited, and if most of the applicants must be placed in some of the jobs the selection procedure comprises. This is the actual position, for example, in the military classification process, where each conscript who fulfils certain minimum requirements of a psychological and physical nature must be placed in some job according to the personnel quotas established. The classification has to be made in such a manner, that the total result, according to the existing productivity criteria, is maximized.

How to maximize the result in accordance with the existing knowledge of the probable level of success of each individual for each possible assignment, is at present under investigation. A number of different attempts have been made to solve this problem (Votaw & Dailey, 1952, and Dwyer, 1954). The main principle is indicated in the following quotation from a report by Votaw & Dailey (1952, p. 1):

> The purpose of the research summarized in this bulletin is to develop techniques of assigning the available personnel to the available jobs in such a way as to maximize the efficiency of utilization of manpower. An important assumption regarding the techniques is that the assigning agency can determine a quantitative index for each possible person–job pairing, representing the person's productivity or pay-off to the service if assigned to that particular job. The manpower will be utilized with maximum efficiency if the assignment of all persons to all jobs is such as to maximize the sum of the pay-off amounts for the finally determined combination of persons and jobs.

If personnel classification by such a method is to be successful, it is not enough for the tests to give good prediction only for each one of the various jobs. As previously mentioned, this can often be done just as well by certain general ability tests, that "skim off the cream". An example is a general classification test of the kind used by the Armed Forces in many countries. Such a test has proved to give good prediction also for most jobs in the Swedish Armed Forces. A personnel classification system which is mainly based on such a test leads to a great demand for personnel with high scores in just this test, and the supply of highly qualified personnel to be drafted in this way will be insufficient. The consequences of this shortage are beginning to be taken into consideration, particularly in the United States, by both military and civilian organizations in their search for high level talent. The problem has been discussed from a psychometric viewpoint for example, by Daily (1951; pp. 17–18) in an article entitled The Supply and Identification of High Level Talent, where he writes:

> The supply of high level talent will be in part a function of the number of relatively independent dimensions of individual differences that are generally accepted as significant talents. ... We can see that if talent is regarded as unidimensional (general intelligence for example) the supply of high level talent will be X percent. However, if talent is regarded as existing in two independent dimensions, then the supply of high level talent becomes X percent plus $X\ (100-X)/100$ percent or very nearly $2\ X$ percent if X is small as would usually be true. Of course, the increase becomes considerably less if the two dimensions are correlated. The supply will be asymptotic to 100 percent as the number of independent dimensions becomes large, and even with moderately correlated dimensions the supply can become relatively large with a reasonably small number of dimensions.

This type of personnel classification, known as differential prediction, must, to a greater extent than direct prediction, be based on abilities unique for each job or job family. The more the selection for each of the jobs can be based on one special combination of tests which is low-correlated with test-combinations for other jobs the greater is the probability that the different jobs will have their personnel quotas filled without competition for the same type of personnel. This presupposes that the tests are valid only for certain occupations and are mutually low-correlated. This is emphasized, for example, by the experts of the U.S. Air Force in this field, who have been very active in the development of the theory and practice of differential prediction. In a report issued by the U.S. Air Force on "Validity of the Airman Classification Battery AC-1" it is stated: (Gragg–Gordon, 1950, p. XIX):

The ACB was developed upon the rationale of differential classification, which implies that each job or group of jobs requires a different pattern of aptitudes. The prediction of success in each of these jobs or groups of jobs requires the use of many tests which can be combined into weighted composites, depending upon their intercorrelations and correlations with the criteria of success on the job. The most efficient classification occurs when the separate tests, each as pure as possible, measure different aptitudes, or correlate low with one another and high with the criterion. The differentiation possible by use of such weighted composites depends on their validity for differential prediction, which, as is indicated above, is a function of differences in validity coefficients and the size of intercorrelations of tests.

An essential requirement for differential prediction is low intercorrelation between tests, since the prediction for each individual is also based on *differences* in aptitudes for the different jobs described, for example, in a profile. This has been shown, by Mollenkopf, who gives a formula for the validity coefficient in prediction of the difference between two criterion scores a and b with the help of the same test battery (Mollenkopf, 1950, p. 414):

$$R^2_{d*d} = \frac{R^2_{a*a} + R^2_{b*b} - 2\,R_{a*a}\,R_{b*b}\,R_{a*b*}}{2\,(1 - r_{ab})}. \tag{66}$$

The notation * indicates predicted number. R_{d*d} is the correlation between the predicted difference and the actual difference; R_{a*a} is the multiple correlation between the battery and the criterion a; R_{b*b} is the multiple correlation of the battery with the criterion b; r_{a*b*} is the correlation between the weighted test scores for prediction of respective criteria and r_{ab} is the correlation between the criteria a and b.

Mollenkopf points out that r_{a*b*} i.e. the correlation between the weighted test scores, is the determining factor for the success of the differential prediction (op. cit., p. 415). This is clearly seen when $R_{a*a} = R_{b*b}$, which gives

$$R^2_{d*d} = R^2_{a*a}\,\frac{(1 - r_{a*b*})}{(1 - r_{ab})}. \tag{67}$$

From this formula it is evident that R_{d*d} approaches unity when r_{a*b*} approaches zero. r_{a*b*}, which is the correlation between the sums of weighted test scores, can also be calculated as a weighted sum of correlations between tests. The correlation r_{a*b*} becomes smaller, the lower the correlations are between the tests, and the greater the differences in weights for the two criteria.

There are several discussions on the requirements of differential prediction in respect of the tests (Mosier, 1951 b, and French, 1955). Consequently it is not necessary here to go more fully into this problem. It has already been shown that in order for differential prediction to be successful, tests are required which are valid only for certain jobs and which have mutually low correlations. Factor tests are supposed to have at least the second property.

The procedure employed in the construction of factor tests for differential prediction has much in common with that which has been discussed in connection with direct prediction. A complication may, however, arise. If all criteria are highly correlated and thus have a similar factor structure, they will be correlated with the same tests, and the possibility of making differential predictions of these criteria will consequently be small. The correlations between the criteria can be high, either because the jobs that they correspond to are actually highly correlated or because the criteria are biased in the same direction. Both these cases are present, for example, in the military personnel classification procedure. In this case, the criteria are often marks based on results of training, which may be expected to be highly correlated. Moreover, there is a tendency to give marks which are based perhaps too much on—for all branches of training—common variables such as general educational ability, adjustment to other conditions in military life, military conduct, and willingness.

Should the criteria be unbiased but still mutually highly correlated, this indicates that the real basis for successful differential prediction is small. Performances within the group of occupations or job families are in regard to prediction, in other words, dependent on the same factors.

An intermediary form between direct and differential prediction is comparative prediction. It is applicable, for example, to vocational guidance where the client wants to know both which jobs that are most suitable for him (differential prediction) and how well he is suited for them (direct prediction). He needs a vocational profile in which the differences between the profile values are valid and where each profile value gives valid prediction of the job in question. French (1955, p. 27) expresses this as follows:

> Where there is needed a decision as to which of two occupations or courses of study offer an individual the greater chance of success, differential prediction provides the direct answer. However, in the usual case, there is not only interest in which course of action is best but also there is interest in how successful each course of action is likely to be.

The requirements of the test are substantially the same, both in comparative and differential prediction. The validity of each test must be limited to certain jobs and the correlations between the tests must be low. The factor test is thus also a possible aid in the solution of the problem of comparative prediction.

Conclusions about Factor Tests as Predictors

Even if it can be maintained that tests based on factor analysis constitute theoretically a rational solution of many prediction problems, the practical results hitherto obtained in the form of statistically reported validities are not particularly numerous or convincing. There are several reasons for this, besides those mentioned above.

The degree of resemblance between the different jobs or job families may be such, that it can be described, to a great extent, by a general factor. A test measuring this general factor gives direct prediction for all jobs, to a relatively high degree, but cannot lead to successful differential prediction, which is of particular value. It is necessary to have a test measuring the specific variance for each job.

Another possibility is, that the requirements for some jobs are so specific in their structure that no common factor of importance can be found. In this case, tests measuring special factors are required both for direct and differential prediction.

Another difficulty is, that it is not always possible to divide the variance in a certain job performance into factors. This is pointed out, for example, by Fleishman, who is engaged in constructing valid, and at the same time, low-correlated psychomotor tests, notably for differential prediction in the U.S. Air Force. He says (Fleishman, 1953, p. 45):

> There is an assumption here that it is possible to break down the variance in performance of complex psychomotor tasks into simpler, more fundamental psychomotor functions. Whether the introduction of complexity of function (e.g. tasks requiring coordination or integration of operations) introduces variance not reproducible by *any number* of more analytical tasks remains a problem for future research.

In many cases the requirements for the different jobs can in themselves be favourable to both direct and differential prediction with the help of factor tests but those criteria for success in the jobs, that can be obtained, are so imperfect, that a rational test construction program with the help of statistical technique cannot be carried out. The criteria have often so low a validity and reliability that the unique parts of the variances dominate. Validation is also impeded if the criteria are biased in the same direction, for example, through the influence of a halo effect. It can cause the appearance of an illegitimate common factor that complicates validation, particularly of tests for differential prediction.

In many cases the criterion of success in a job is complex and composed of a number of subcriteria, which should be given weights in relation to each other. The technique of canonical correlation (Hotelling, 1935) can be used for finding the set of weights for subcriteria and tests, which gives the maximum correlation between the two groups of variables. Mostly it is, however, more relevant to give weights to the different subcriteria on non-statistical grounds, e.g. educational or psychological considerations. Peel (1948) has dealt with this problem, and has developed a method for maximizing the correlation between a set of subcriteria, given certain arbitrary weights, and a set of tests. The problem of finding the arbitrary weights for the subcriteria is rather neglected. More attention has recently been paid to this problem (Ghisselli, 1956 and Sluckin 1956). A conceivable way of finding the dimensionality of a criterion would be to factor analyse the different parts of the criterion, with or without the use of reference tests. This should give a picture of the criterion structure and should also contribute to a reduction of the number of subcriteria to a smaller set.

Even if the criteria are unbiased and of high validity and are in other respect such that the possibilities for differential prediction are good, the technical difficulties to produce differential prediction are very great. Generally, each individual tested, will get a criterion score only in the job for which he has been selected. The theory of differential prediction—for example, formula (66)—presupposes, however, that each individual has scores in all criteria. The theory works with a matrix including all correlations between the different tests and criteria for the whole sample of individuals who are tested with the battery. In practice it is not generally possible to get correlations between the different criteria. Furthermore, the correlations between the tests and criteria will be based only on the sub-sample selected for each particular job.

The statistical treatment can also be complicated by restriction of range, caused by the fact that a test which is to be validated, has also been used in the selection for the job. If this restriction of range is of the uncontrolled type based on other variables, then the common correction formulae are not directly applicable (Gulliksen, 1950, p. 162).

It is one thing to establish the factor structure for a job criterion, another to work out the tests that measure these factors in a pure form. Factor tests, particularly those which are useful from the viewpoint of time and cost, tend to be impure and have a much higher correlation with one another than corresponding factors. The tests of Primary Mental Abilities, constructed by Thurstone on the basis of his factor analytical investigations, for example, have these shortcomings. This is evident from correlations between the tests of Primary Mental Abilities published, for example, by Bond & Clymer (1955).

The result of these and other difficulties is, that, at present, there are rather few consistent examples of how factor tests have been used successfully in connection with prediction, especially differential prediction. Particularly in the United States, however, relative optimism is felt concerning the possibilities of achieving this. An example of this is the quotation from French on page 64. In other countries, especially in England, a more critical attitude is often adopted (Vernon, 1950a, p. 128). This is also the impression one gets from available investigations from different countries. At present, it seems as if factor analysis has mainly contributed to the development of the theory of differential prediction. Practical results are, to a great extent, lacking (Terman, 1954, pp. 225–226, and Cronbach, 1956, p. 178).

The Problem of Finding Explanatory Factors.

A factor analysis of a set of data may be regarded as a deductive procedure. The different series of observations are combined with each other to form a correlation matrix, which is then analyzed; the whole procedure is conducted according to certain mathematical rules. The factor loadings obtained are deduced from the observed values. This procedure, in itself, only gives the factors descriptive properties. If, besides this, they are to have an explanatory meaning, then the analytical method or the further steps which follow must be so constituted, that the factors obtain also some new characteristics typical of explanatory concepts. Such characteristics are, for example, simplicity, invariance, and conformity to results obtained by other methods of observation and inference.

The meaning of explanatory factor analysis is unclear. This is due in part to the fact that much in the terminology, e.g. the factor concept itself, is ambiguous and used in varying contexts. Furthermore, statements made by different scientists on scientific theory very often imply differences of opinion. Both in psychology and in the general theory of science, these questions are still being debated. In the next chapters an attempt will be made to determine, on the basis of this debate, what meaning can be given to explanatory factor analysis, and whether it is possible to conduct such analyses, and what methods should then be applied.

CHAPTER IV

FACTOR ANALYSIS AND THEORY OF SCIENCE

In the behavioral sciences there is a need for a firmer foundation for research methods and for development of scientific theories and laws. Some of the attempts made to build a foundation are known as the theory of science or the philosophy of science or the science of science, and are greatly influenced by modern philosophy, for example by the school of logical empiricism. In the following pages, an attempt will be made to utilize some of the results of the theory of science, which are relevant to a discussion on the meaning of, and the distinction between, descriptive and explanatory factors. Even if much of the following discussion may prove to be one-sided and oversimplified, we still hope that it may nevertheless have a constructive contribution to make, by arranging the problems, and suggest solutions, in a way which will enable subsequent investigators to solve them.

One of the typical features of the theory of science is, that it takes the methods and results of empirical investigation as its starting point, and then endeavours to clarify their meaning and implications. Hence, it is not mainly an isolated attempt to advance cosmological theories, in the form of general theories of nature, but a concentrated endeavour to analyze the methods and theories which are developing in the different sciences. Of particular interest in this connection are the logico—semantic analyses of scientific theories and their elements.

Two closely related fields of investigation will now be considered. They both have an intimate bearing on the meaning of, and the distinction between, descriptive and explanatory factors. The first, is the general debate, within the theory of science, on the distinction between descriptive and explanatory models, and the second is the discussion carried out mainly in the *Psychological Review*, concerning intervening variables and hypothetical constructs.

Description and Explanation in the Theory of Science

One of the tasks of science is to establish laws governing empirical facts. These laws should connect observations of separate facts into a logical system, making possible the communication of knowledge and prediction of new facts and events (Braithwaite, 1953, p. 1). The ideal would be if all established regularities and other facts could be deduced, as logical consequences, from a few universal laws. A system of such laws constitute a scientific theory that is said to be an explanation of the empirical facts. A scientific theory thus consists of a number of terms linked together to propositions or formulae through a language of logical or mathematical nature, in such a way that the empirical facts can be deduced from these terms. These formulae and propositions should be of such a nature that they can form a basis for hypotheses which can be empirically confirmed.

To avoid "metaphysical" infusion into the theories, some scientists, e.g. among the operationists, maintain that scientific theories ought to be based on concepts and laws which are deduced entirely from facts empirically established (Marx, 1951). This opinion has been criticized by e.g. Carnap (1956) and Hempel (1952). The latter says about theories deduced from empirical data (p. 23–24):

> The conjecture mentioned in the preceding paragraph may now be restated thus: Any term in the vocabulary of empirical science is definable by means of observation terms; i.e. it is possible to carry out a rational reconstruction of the language of science in such a way that all primitive terms are observation terms and all other terms are defined by means of them. This view is characteristic of the earlier forms of positivism and empiricism, and we shall call it the *narrower thesis of empiricism*. According to it, any scientific statement, however abstract, could be transformed, by virtue of the definitions of its constituent technical terms, into an equivalent statement couched exclusively in observational terms: Science would really deal solely with observables. ... Despite its apparent plausibility, the narrower empirist thesis does not stand up under closer scrutiny.

The criticism against "the narrower thesis of empiricism" points out that explanations based on observational terms have a tautological character, since they do not add any new information. A scientific theory of this nature is primarily a translation into a suitable language or formula system of empirical data and can give a summary and outline but, as a matter of principle, no new information. Such a theory is only an alternative description of observed facts in a condensed and more convenient form.

According to Hempel (1952, pp. 23–39), analyses of existing scientific theories show that many of them are based on theoretical constructs, the content of which is not exhausted through deduction from empirical facts. He considers such elements in a theory necessary in order to make the theory simpler and easier to develop. Such theoretical constructs are not fully confined to their content through deductions from empirical data and have openness of meaning. They can thereby be endowed with qualities which make it possible to link them to other theories.

A scientific theory should, in Hempel's opinion, state both more and less than the empirical data on which it is based and which it is intended to explain. It gives more information by including the empirical observations in a larger system, and less information by limiting itself to the essential aspects of the data by the exclusion of details. Such a theory can be further developed because it is built up of concepts and sentences which derive their content from laws and theories established in other areas. Hempel expresses this idea as follows (p. 36–37):

> The entire history of scientific endeavour appears to show that in our world comprehensive, simple, and dependable principles for the explanation and prediction of observable phenomena cannot be obtained by merely summarizing and inductively generalizing observational findings. A hypothetico-deductive-observational procedure is called for and is followed in the more advanced branches of empirical science: Guided by his knowledge of observational data, the scientist has to invent a set of concepts—theoretical constructs, which lack immediate experimental significance, a system of hypotheses couched in terms of them, and an interpretation for the resulting theoretical network; and all this in a manner which will establish explanatory and predictive connections between the data of direct observation.

The main question is whether there is any difference in principle between explanation and description. Here we have limited "explanation" to mean that the special facts can be deduced from general concepts. This general qualification is found, for instance, in a paper on the meaning of scientific explanation written by Feigl (1953). In this paper the meaning of description is also discussed. This is illustrated by the following quotation (p. 511):

> It is very helpful to restrict the meaning of "description" to singular statements representing fully specific facts, events or situations. Such descriptions may appear as the conclusion of explanatory, i.e. deductive inferences. Some of the premises of these inferences must then be scientific laws or theoretical assumptions. Since laws and theoretical assumptions are (or at least contain) generalized statements (i.e. unlimited universal propositions) they are not classified as descriptions.

These distinctions between explanation and description are probably generally acceptable but there are different opinions as to what should be included in the concept of "generalized statement" in the definition of explanation. If the generalized sentences and concepts, included in what are explanations from a logical viewpoint, were only deductions from "singular statements representing fully specific facts", then it is possible to concede that in principle there is no difference between explanation and description. It would then be only a difference in degree, where explanation on the whole referred to more "singular statements, representing fully specific facts", than description. On the other hand, as previously mentioned, some scientists maintain that at least some of the propositions and terms included in the explanations are theoretical constructs with openness of meaning, and make possible the formulation of other laws and theories (Carnap, 1956, pp. 66–69). This assigns to these constructs a surplus meaning over and above that which can be deduced from the specific facts. In principle, this introduces something new at least into certain explanations, besides what is included in description.

The next question which has partly been touched upon, is concerned with what characteristics an explanatory concept should have, in order to be acceptable and useful. The requirements which are mentioned in the following pages, and which are taken from the field of the theory of science partly overlap. We shall not try here to separate them into logically independent concepts, or to determine which are the most important.

Explanatory concepts ought to have the following characteristics:

(1) *Comprehensiveness.*—Which means that the concept can be used in the deduction of many empirically established regularities and relations. For example, the concept of wave in physics brings together a number of optical, acoustic, and electrical phenomena under one common theory.

(2) *Openness of meaning.*—This is related to the requirement for comprehensiveness, and emphasizes that the concepts ought to be capable of development, so that new testable hypotheses can be deduced from the concepts and new facts fitted into still more general theories.

It is mentioned above, that, for example, Braithwaite and Hempel have tried to show that only theoretical constructs are capable of development because they have surplus meaning. A theoretical construct however, is not only a metaphysical concept, since it is connected to empirical data by epistemic correlations which make confirmation possible (Northrop, 1947, pp. 119–131).

(3) *Simplicity.*—This requirement is often emphasized in scientific theory construction. The notion of simplicity implies that a theory should be made up of as few logically independent propositions and theorems as possible (Carlsson, 1949, pp. 132–143). Each term must therefore include a large number of phenomena and form an integral part of many theories.

(4) *Completeness.*—An explanation also ought to give a correct and rather full description of the facts that are to be explained. This may conflict with the requirement for simplicity and may make it necessary to revise explanatory concepts when new and more exact data become available.

(5) *High degree of confirmation.*—This signifies that it should be possible to deduce a great number of empirically confirmable conditions from the explanatory concepts. These confirmable conditions ought to be such that, in principle, the concepts can

be verified or disproved and may also be compared with alternative explanatory concepts.

In many connections, it is required that explanatory concepts should have "reality" or "meaning". These terms readily acquire a metaphysical tinge, and in the theory of science there is a marked tendency to replace them by the requirement for confirmation (Carnap, 1956).

(6) *Invariance.*—A useful concept should be derivable and confirmable in many independent ways. This leads to the requirement that the content of the concept must be the same in different connections and in various empirical systems.

Intervening Variables and Hypothetical Constructs

The current trend in psychology is characterized by an endeavour to link up different, empirically established, facts into strict psychological theories. The works by Tolman (1932) and Hull (1943) contain examples of such constructions of theory. In these and similar publications there is a definite tendency noticeable to give the theories, as far as possible, an operationally definable content. At the same time, theories should be simple and clear, which means that the variables should be few and fundamental in character, each comprising a number of important conditions for the behaviour studied. Tolman termed such variables intervening variables and required that they should be deducable from operationally definable conditions (Tolman, 1936).

Difficulties were experienced in connection with these attempts. Concepts not fully reducible in accordance with operationally definable conditions were sometimes included in the theories. Nor was the meaning of the concept intervening variable made fully clear. Hull, for example, introduced intervening variables in his theory in such a way that they sometimes seemed to be synonyms of what he called "hypothetical entities" (Hull, 1943, pp. 21–22).

MacCorquodale and Meehl (1948) have attempted to clarify this question in a paper in the *Psychological Review,* "On a distinction between hypothetical constructs and intervening variables". In this paper they tried to limit intervening variables to the operationally definable sphere and to give other properties to the concept hypothetical construct. Their aims are clearly stated in the summary of the paper, as follows:

> 1. At present the phrases "intervening variable" and "hypothetical constructs" are often used interchangeably, and theoretical discourse often fails to distinguish what we believe are two rather different notions. We suggest that a failure to separate these leads to fundamental confusions. The distinction is between constructs which merely abstracts the empirical relationships (Tolman's original intervening variables) and those constructs which are "hypothetical" (i.e., involve the supposition of entities or processes not among the observed).
>
> 2. Concepts of the first sort seem to be identifiable by three characteristics. First, the statement of such concepts does not contain any words which are not reducible to be empirical laws. Second, the validity of the empirical laws is both necessary and sufficient for the "correctness" of the statement about the concepts. Third, the quantitative expression of the concept can be obtained without mediate inference by suitable groupings of terms in the quantitative empirical laws.
>
> 3. Concepts of the second sort involve words not wholly reducible to the words in the empirical laws; the validity of the empirical laws is not sufficient condition for the truth of the concept, inasmuch as it contains surplus meaning; and the quantitative form of the concept is not obtainable simply by grouping empirical terms and functions.
>
> 4. We propose a linguistic convention in the interest of clarity; that the phrase *intervening variable* be restricted to concepts of the first kind, in harmony with Tolman's original definition; and that the phrase *hypothetical construct* be used for those of the second kind.
>
> 5. It is suggested that the only rule for proper intervening variables is that of convenience, since they have no factual content surplus to the empirical functions they serve to summarize.
>
> 6. In the case of hypothetical constructs, they have a cognitive, factual reference in addition to the empirical data which constitute their support. Hence, they ought to be held to a more stringent requirement in so far as our interests are theoretical. Their actual existence should be compatible with general knowledge and particularly with whatever relevant knowledge exists at the next lower level on the explanatory hierarchy.

Discussions have been conducted along these lines. Several contributors, for example, Bergmann (1953) and Maze (1954) doubt if it is really possible to maintain the distinction between intervening variables and hypothetical constructs and if it is a fruitful one in scientific theory construction. Bergmann (1953, pp. 445–446), for example, emphasizes that as soon as intervening variables are included in a theory, they automatically acquire both surplus meaning and the status of hypothetical constructs. This criticism has been dealt with for example, by Rozeboom (1956), who has tried to define more precisely the meaning of intervening variables and hypothetical constructs, and the distinction between them. He emphasizes that an intervening variable is derived from certain empirical data through an analytical relation (p. 253):

> An intervening variable (transformation variable) is a variable which has been systematically *defined* in terms of its antecedents and is hence dependent upon the latter for its meaning. To recapitulate the analysis by means of a formal example, let us suppose that $\hat{z} = (x + y^2)/w$ is an empirical relation expressing the best estimate of a known variable z as a joint function of other known variables w, x and y. We are now able to express $\hat{z}$ as a function of analytical mediation variables by defining one or more of these as mathematical functions of w, x and y.

Such an intervening variable, can, if it is appropriately chosen, combine certain empirical data into useful and operationally definable concepts. These concepts do not have any surplus content besides that which is derived from the empirical relation they are based on. Hypothetical constructs on the other hand, have such a surplus content, and are delimited in the following way (op. cit., pp. 254–255):

> *Hypothetical constructs.*—In the last section we considered those mediation variables which derive analytically from their antecedents. For the remainder, passage from antecedents to mediators must be made with less than logical certainty, and such variables we shall designate as "inferred variables". Precisely because inferred variables are not analytically contained in their antecedents, it follows immediately that (*a*) the meaning, or ontological content, of the inferred variables is not reducible to that of its antecedents, and (*b*) a certain degree of error must be expected in passage to the inferred variable from its antecedents.

Roseboom thinks that the notation "hypothetical" for these constructs has overemphasized their unverified nature. Many concepts, which are used in scientific theory and which are well confirmed, are, in his opinion, still to be considered as hypothetical constructs in principle (op. cit., p. 255):

> I suggest that by "hypothetical construct" we understand all cases of inferred mediation variables, even though in a few instances the generic existence of all the variables and its applicability to the objects concerned may admit of little practical doubt.

A similar position is taken by Ginsberg (1954). He especially emphasizes that concepts relevant to areas other than those where they are operationally defined, aquire the status of hypothetical constructs.

An operationally oriented scientist like Marx (1951) accepts hypothetical constructs as useful tools in the process of building a psychological theory, but he considers them most useful in the preliminary stages of scientific research in a special field. In this connection, they can give rise to fruitful hypotheses for empirical investigations, and thus be converted into operationally valid intervening variables. In his opinion, only operationally definable concepts are finally acceptable in a sound scientific theory.

The main impression gained from reading the literature on the subject is, that hardly any scientist denies that hypothetical constructs have their part to play as a link in psychological theory construction, and even Tolman (1951, p. 282) has accepted them. Here, we do not have to determine to what extent it is both possible and desirable to try to reduce them completely to operationally definable concepts. However, even if one does not wish to distinguish between intervening variables and hypothetical constructs, one nevertheless has to accept the existence of both of some kind of descriptive concepts which are operationally definable and theoretical concepts with surplus meaning.

It is difficult to summarize the discussion, which apparently is far from finished. If intervening variables, or whatever we may choose to call them, are considered as variables which are fully reducible from an empirical relation, it would appear that these concepts are insufficient for the construction of scientific theory. Concepts which are not fully operationally defined are also required. This requirement is satisfied by hypothetical constructs which have the comprehensiveness and openness of meaning, which make them useful in scientific theories that cover a number of regularities obtained empirically and that form a basis for hypotheses as to new theories. Scientific theories, which work with such hypothetical constructs, have, in most sciences, been very fruitful also for further empirical research.

The Distinction between Descriptive and Explanatory Factor Analysis

The review of the discussion on the theory of science indicates that it is fruitful to make a distinction between descriptive concepts deducible only from empirical data, and, theoretical constructs with surplus content and openness of meaning. The latter enable us to formulate a number of propositions and theories, which can be confirmed by many different methods on a wide range of data.

Accordingly, such factors should be called descriptive, as are operationally defined by the use of a specific method employed in a specific test battery and administered to a specific population. Those factors may be called explanatory which have surplus content referring them to factors found in other batteries and populations. A number of other confirmable propositions and theories can be deduced from this type of factors.

Hartley (1954) has tried to make a distinction between descriptive, and what he calls inferential factor analysis, and has discussed how the latter should be conducted. He delimits descriptive factor analysis in the following manner (op. cit., pp. 196–197):

> Thus, by descriptive factor analysis is herein meant the kind of factor analysis that aims at converting the correlation matrix to a more usable form. The facts the descriptive factorist wishes to deal with are represented solely by the correlation matrix. His factor analysis, like the draftsman's drawing, adds no factual material but contents itself with an exact restatement of the factual material expressed by the correlation matrix. Hence terms like "right", "wrong", "error", "true", "false", are inapplicable to the factors of descriptive factor analysis.

There are factor studies which, besides a purely descriptive presentation, attempt to give properties to the factors, referring to such concepts as "underlying unities". Hartley gives the following definition of this type of factor analysis (op. cit., p. 197):

> It may be possible, then, to characterize adequately inferential factor analysis as that form of factor analysis in which the factor matrix has empirical meaning in addition to that expressed by the correlation matrix. Two different factorial solutions for the same correlation matrix could, then, express different and even contradictory meanings. The meaning of the factor matrix which is in addition to that expressed by the correlation matrix is the reference to the unobserved but hypothesized "factors". Since these factors are not referred to by the correlation matrix, the correlation matrix would seem to be serving as the evidence from which the factors are inferred by some process or other. Hence the term "inferential" has been selected to characterize this form of factor analysis.

When we wish to define our attitude towards the question whether inferential factor analysis is possible, it is, according to Hartley, first necessary to find out if such factor analytical terms as "underlying unities" have any definite meaning. Hartley deals with this problem in his unpublished doctoral dissertation (Hartley, 1952). After making a logical analysis, Hartley finds that it is possible to formulate a language which gives such properties to the factors. This language covers also new properties of the factors besides those deduced from the data in a specific investigation. For factors which possess inferred properties it is significant to ask whether they are "right" or "wrong".

When Hartley discusses criteria for judging whether the factors are "right", he refers to Carnap's discussion of scientific theories. Carnap (1950) points out that an important criterion of the validity of scientific theories is their degree of confirmation. The degree of confirmation of a theory is higher, if it can be verified by different empiricial investigations and if it can form the basis for new testable propositions. From this point of view, Hartley (1952, p. 110) gives the following criterion:

> ... inferential factor analysis as herein conceived, should accept as its proper solution, the solution having the highest degree of confirmation on the evidence present.

One way of obtaining a higher degree of confirmation is to make several factor studies of samples from different populations and different tests. This is the same as making studies of invariance.

Hartley believes that inferential factor analysis, based on the principle of simple structure, has a higher degree of confirmation than others types of factor analysis. This is, because it is based on the principle of parsimony. He finds it difficult to show this, however, because the principle of simple structure is so vaguely formulated.

Hartley's distinction between descriptive and inferential factor analysis seems fruitful in the light of the discussions on the theory of science which have been reviewed above.

When we try to interpret the factors found through factor analysis and to give them an empirical basis, which makes the range of their validity wider than that of a certain constellation of tests and a certain population, and which includes them in a more general psychological theory, then we provide them with surplus meaning and thus convert them into hypothetical constructs. Factors with such properties are useful as explanatory concepts in psychological theory, since they are obtained by different factor analyses and various kinds of investigations of different types of tests and populations. A consequence of this is, that explanatory factors should have the property of being invariant when changes in the tests and populations are made.

In summing up, it may be stated that it is both possible and useful to make a distinction between descriptive and explanatory factor analysis. Descriptive factor analysis is limited to obtaining factors, which, from a certain point of view, are useful as descriptions of a certain constellation of data. Explanatory analysis is a link in an inferential procedure, whose object is to find factors that explain as many regularities and relations as possible, even outside of an actual factorial study.

CHAPTER V

EXPLANATORY FACTOR ANALYSIS

The Testing and the Generating of Hypotheses

We may distinguish between two ways of using factor analysis in an inferential procedure for finding explanatory factors, namely, for testing hypotheses and for generating hypotheses (Eysenck, 1953).

As has already been shown in the historical survey in Chapter II, psychologists, such as Spearman, Holzinger, Burt, and Thomson, mainly limit their use of factor analysis to the testing of hypotheses already formulated. They employ a procedure where one or more factor analyses, together with other information, will show whether hypotheses, previously advanced, as to the nature of the underlying processes can be confirmed or not.

In such a procedure, hypotheses as to structure, are tested on various sets of data. If the factor structure cannot be applied to the data, the hypothesis can be rejected. If, on the other hand, the analysis supports the hypothesis, this is an indication of the correctness of the hypothesis. Conformity, however, is only to be considered as a necessary, and not as a sufficient, basis for the correctness of the hypotheses, since many other hypotheses as to factorial structure may also prove to be concordant with the nature of the data. This type of factor analysis, like the testing of hypotheses in general, has the drawback that a hypothesis can be rejected but it cannot be proved true and thus enable us to exclude other hypotheses. The only thing which can be safely said, when empirical data are found to support a hypothesis, is, that the hypothesis gives one of many possible descriptions of data. If the present hypothesis as to the factor structure, should also be advanced to explain why the relations between the observation series have certain properties, then also other requirements for the validity of hypotheses must be induced. These requirements ought to be such, that as many alternative hypotheses as possible are rejected. This necessitates factor analyses of various populations and with differently composed test batteries and even entails investigations or considerations other than those of a purely factor analytical nature.

The testing of hypotheses by means of factor analysis can also be combined with procedures that differ from those which vary population or test battery. One can try to study, for example, changes in the supposed factor by experimental variation, growth and training. The influence the factor is exposed to, should then change its loadings in the same proportion in all the variables where the factor is included, while the loadings of the other factors remain unchanged. For example, in a brain-surgery operation that reduced the variance of a memory factor by half in one test, the loadings of this factor should also be reduced proportionally in other tests.

The fewer, the hypotheses which are tested together at the same time, the safer their confirmation, as a rule. It is often possible to get the results of a factor analysis to correspond to those required by certain of the hypotheses, if we capitalize on errors of sampling and measurement. The observed variables should consequently be selected in such a way, that the principal features in the common variance of the battery, which are not being studied, are known through earlier results. If part of the structure is already well established in this way, one can study with more confidence whether the hypothesized factors are really found in the tests, where they should be loaded according to the hypotheses.

When operational hypotheses are tested by factor analysis, this forms only a link in investigations, which should include other studies, for example, of a clinical, genetical, or experimental nature. The greater the number of operational hypotheses which can be advanced and verified within the limits of a main hypothesis as to an underlying factor, the more reason there will be for maintaining that the factor is "fundamental" and indicates some basic and universal characteristic of the underlying structure.

To summarize, it may be said that hypotheses can be tested by means of factor analysis. Whether the factors that pass such a hypothesis-testing are to be given an explanatory meaning or not, will then also depend even on the extent to which other evidence exists that the factors possess explanatory characteristics.

It is not possible to maintain any clear distinction between testing and generating of hypotheses. By hypothesis-testing some of the hypotheses can be disproved, and at the same time a basis can be found for establishing of new ones. These, in their turn, require new tests, and so on. Nor is the generating of hypotheses a completely isolated procedure. It is usually based on some form of more or less implicit hypotheses, which become evident in the selection of variables and in the choice of the type of solution. In any case, the factor analyst must work with the hypothesis that some form of structure exists, and that this has certain characteristics, in regard to its general type, which make a unique solution possible.

Hypothesis-Generating Factor Analysis and the Principle of Parsimony

On the whole, it is only Thurstone and his followers who make a serious attempt to assign to factor analysis a hypothesis-generating role. Consequently, the discussion on the possibilities of generating hypotheses by means of factor analysis may be considered in connection with Thurstone's approach. His ambitions are evident from the following statement (Thurstone, 1947, p. 55):

> When a particular domain is to be investigated by means of individual differences, one can proceed in one of two ways. One can invent a hypothesis regarding the processes that underlie the individual differences, and one can then set up a factorial experiment, or a more direct laboratory experiment, to test the hypothesis. If no promising hypothesis is available, one can represent the domain as adequately as possible in terms of a set of measurements of numerical indices and proceed with a factorial experiment. The analysis might reveal an underlying order which would be of great assistance in formulating the scien-

> tific concepts covering the particular domain. In the first case we start with a hypothesis that determines the nature of the measurements that enter into the factorial analysis. In the second case we start with no hypothesis, but we proceed, instead, with a set of measurements or indices that cover the domain, hoping to discover in the factorial analysis the nature of the underlying order. It is this latter application of the factorial methods that is sometimes referred to as an attempt to lift ourselves by our own boot straps, because the underlying order in a domain cannot be discovered without first postulating it in the form of a hypothesis. This is probably the characteristic of factor analysis that gives it some interest as general scientific method.

This latter type of factor analysis, also called explorative factor analysis, is to be used primarily in the mapping of a field about which we have little knowledge or developed theories. The results of such analysis can then be used for formation of more rigorous hypotheses and in planning experiments.

As is pointed out in Chapter II, Thurstone's intentions should be seen against the background of his pragmatical attitude to science. He starts out from the fact that alternative descriptive systems are possible and considers that in the choice between these, the principle of parcimony should be decisive. This principle should then be applied within the whole field that the scientific theory is to cover and to all empirical facts that can be deduced from the theory. This broad application of the principle of parcimony has important consequences. It means, for example, that the number of concepts should be few and invariant in regard to their content in all their connections. Furthermore, simplicity requires that these concepts shall be applicable in as many connections as possible. The mathematical methods used by Thurstone to find factors satisfying these requirements are also based on the principle of parsimony (Thurstone, 1947, p. 333). This means that the number of common factors used for representing the correlation matrix should be as few as possible. Moreover, each factor should be loaded only in a few tests according to certain rules.

Thurstone wants to find, by means of the principle of parsimony, a few invariant factors which are comprehensive in the development of scientific theory. By applying the principle of parsimony, Thurstone considers that he has avoided metaphysically oriented discussions whether the factors are "real", "existing", etc.

Even in explanatory factor analysis it is necessary to be restrictive in the selection of the problem. This can be done by selecting the tests so that large parts of the factor structure of the battery are already known and by making the number of tests large in comparison with the number of factors in the battery. The aim is to make the factor pattern as overdetermined as possible. This can be achieved if each factor gets loadings in three or possibly more tests, at the same time as the requirement for simple structure is fulfilled. Consequently the selection of the tests should be based on a rough idea of how many factors have to be dealt with, and in which directions they can probably be localized. Thurstone therefore emphasizes, that it is not possible to find a meaningful structure in any arbitrary collection of tests (op. cit., p. 313):

> In planning a factorial analysis it is desirable to cover the domain to be investigated as completely as possible with a large number of related variables, which should be experimentally independent but correlated and representative of different aspects of the domain. When a study has been planned this way, it will nearly always happen that every test vector lies near some other vector in the system. This arrangement is desirable for several reasons, principally in order to make it possible to discover a simple structure if it exists in the domain.

Whether, and to what extent, there is any simple structure in the field is an open question, which the analysis will have to answer. In an explorative study of a completely new field it often happens that only a part of the structure is overdetermined and that in a more or less obscure way. But this, however, can form the basis for the design of new factor analytical studies or for investigations of a different nature, that lead to a safer determination of the factors and their characteristics. In this way explorative investigations lead, step by step, over to the testing of hypotheses. Factor analysis, according to the principle of simple structure, is intended to afford material for the formulation of a number of more strictly defined hypotheses as to parsimonious, unique, and invariant structures which should then be further tested and modified. Thurstone thinks that such a gradual procedure is the proper way of obtaining factors representing essential aspects of the underlying structure, since the procedure is based on the principle of parsimony, which has been successful in other sciences. On this point, views differ, and both the theoretical and mathematical bases of the simple structure solution have been called in question. Thomson (1948) and Reyburn & Raath (1949) have, for example, discussed these problems critically.

CHAPTER VI

FACTOR ANALYSIS ACCORDING TO THE PRINCIPLES OF SIMPLE STRUCTURE

The Meaning of Simple Structure

The principles of simple structure are fundamental for Thurstone's methods and make it possible to find factors with a range beyond pure description. The principles of simple structure have been advanced to meet the requirements for parsimony, invariance, uniqueness, and functional unity, which have been previously mentioned.

Restrictive assumptions must be made in order to get a unique solution. The series of restrictions included in the principles of simple structure are based mainly on the principle of parsimony. In this connection, the latter principle means that every test should have as simple a factor pattern as possible, and conversely, that the estimation of each factor should need a combination of only a few of the tests included in the battery. These requirements must be stated more precisely to make a unique and invariant solution possible. Thurstone tries to do this in the following way (op. cit., p. 320):

> If an analysis is to be made by the principle of simple structure, then the investigator gambles that the complexity of each test or measurement is less than the complexity r of the battery as a whole. He hopes that the test complexities are considerably less than that of the battery. If the test complexities are less than r, then there will be one or more blank cells in each row of the factor pattern. Such a pattern will be considered as evidence of an underlying simple order in which each parameter or process in the domain is not involved in every test measurement.

The above requirement of a complexity in each variable, which is less than the rank r in the matrix R, makes a more strict formulation of the principle of simple structure possible. Its meaning can be illustrated by the vector theory, mentioned in Chapter I. The different tests included in a test-battery have their counterpart in test vectors. Since the factor analysis only applies to the common part of each of the n number of tests, each test vector will have a length equivalent to h, i.e. the square root of the communality h_2. The correlation between the tests can be expressed as functions of the angles between the test vectors and of their length, according to the formula

$$r_{jt} = h_j h_t \cos \phi_{jt} \quad (j, t = 1, 2, \ldots, n;\ j \neq t).$$

ϕ_{jt} is the angle between the vectors h_j and h_t. This theory is given in more detail by Holzinger & Harman (1941).

The test vectors in a test battery form a cluster of vectors, with certain angle relations to one another in a space with as many dimensions as the rank of the correlation matrix. The aim of the analysis is to refer this cluster of test vectors back to a system of reference vectors that fulfils the requirement for simple structure. According to the requirements for such a structure none of the tests should be loaded with all the primary factors. In terms of vectors this means that the reference vectors should be chosen so that each test vectors has zero-projection on at least some of these reference vectors.

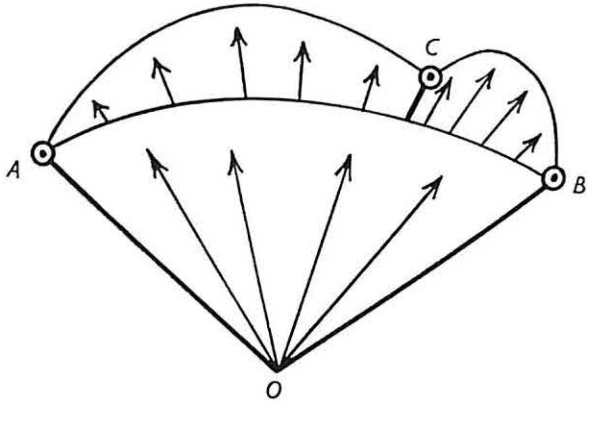

Fig. 4.

This can be graphically illustrated with three factors. The test vectors are indicated in Fig. 4 by a cluster of arrows starting from 0 and forming three fan-shaped planes *AOB*, *BOC*, and *AOC*.

The three planes are defined by three reference vectors Λ_x, Λ_y, and Λ_z, which are normals, i.e., form right angles to each of the planes. The plane *AOB*, for example, includes the test vectors 1, 2, 3, and 4, which are at right angles to the reference vector Λ_x and thus have zero-projection on it, but show certain given projections on Λ_y and Λ_z. The test vectors in the plane *BOC* have zero-projections on the reference vector Λ_y and the test vectors in the plane *AOC* on the reference vector Λ_z. The projections of the test vectors to the different reference vectors correspond to the required factor loadings. In this example, the factors will have loadings according to the following table.

TABLE 4.

Test	Factor		
	x	y	z
1	0	x	x
2	0	x	x
3	0	x	x
4	0	x	x
5	x	0	x
6	x	0	x
7	x	0	x
8	x	0	x
9	x	x	0
10	x	x	0
11	x	x	0
12	x	x	0
13	x	x	0
A	0	x	0
B	0	0	x
C	x	0	0

A test vector in the intersection of two planes will have zero-projection on the two reference vectors forming normals to this plane. This test vector will get projections only on the third reference vector and will consequently measure only one factor. The test vectors *AO*, *BO*, and *CO* are characterized by the fact that each of them is a measure of only one factor. When standardized in such a way that the communality is equal to one, they are called primary vectors, which, under certain circumstances, can refer to primary factors with explanatory properties.

As a rule, the test vectors incorporate r dimensions; and thus r reference vectors are needed. The two-dimensional planes, in the three-dimensional case, correspond to what are known as hyper-planes of r–1 dimensions in the r-dimensional case.

If r is greater than three, the vector system cannot be illustrated in the same way as in Figure 4. Multi-dimensional geometry and matrix algebra can be used, however, in cases where the rank is higher than three.

In carrying out a factor-analytical investigation, we have to start with a cluster of test vectors defined by the correlation matrix. The primary vectors are the unknowns that are to be determined. The purpose of the analysis, according to the principle of simple structure, is to find a set of hyperplanes, each of the dimensionality r–1, which at the same time include many test vectors. Thurstone formulates this as follows (op. cit., p. 328):

> The r linearly independent reference vectors define the same number of co-ordinate hyperplanes. Each hyperplane is a subspace of dimensionality (r–1), and it is defined by a reference vector. The subspace of (r–1) dimensions which is orthogonal to the reference vector Λ_p will be called the co-ordinate hyperplane L_p. If a reference frame can be found such that each test vector is contained on one or more of the r co-ordinate hyperplanes, then the combined frame and configuration is called a simple structure.

Whether a factor model, that meets the requirements for simple structure, is applicable in a certain domain must be decided in each individual case, but Thurstone considers the principles to be so general, that they ought to have wide applicability (op. cit., p. 58):

> Our work in the factorial study of the human mind rests on the assumption that mind represents a dynamic system which can eventually

> be understood in terms of a finite number of parameters. We have assumed, further, that all these parameters, or groups of parameters, are not involved in the individual differences of every kind of mental task. Just as we take it for granted that the individual differences in visual acuity are not involved in pitch discrimination, so we assume that in intellectual tasks some mental or cortical functions are not involved in every task. This is the principle of "simple structure" or "simple configuration" in the underlying order for any given set of attributes.

No one has been able to show how the requirements for simple structure should be formulated in order to give a unique solution. In the absence of a strict mathematical solution of this problem, Thurstone has, besides the minimizing of the total number of factors, proposed the following five criteria, which he has found by experience to give a good basis for the establishment of unique simple structure (op. cit., p. 335):

> We shall describe five useful criteria by which the r reference vectors can be determined. These are as follows:
>
> (1) Each row of the oblique factor matrix V should have at least one zero.
>
> (2) For each column p of the factor matrix V there should be a distinct set of r linearly independent tests whose factor loadings v_{jp} are zero.
>
> (3) For every pair of columns of V there should be several tests whose entries v_{jp} vanish in one column but not in the other.
>
> (4) For every pair of columns of V, a large proportion of the tests should have zero entries in both columns. This applies to factor problems with four or five or more common factors.
>
> (5) For every pair of columns there should preferably be only a small number of tests with non-vanishing entries in both columns.

That these five criteria are sufficient to give a unique solution has not been proved mathematically. On the contrary Reiersöl (1950) has shown that they are not always sufficient, and has tried to formulate the requirements for a unique solution. Probably the five requirements are very often not sufficient for excluding alternative solutions in favour of a specific solution. Thus, Bargmann (1951) has found that in many cases a number of alternative hyperplanes may exist. What can probably be done at present, is to determine the structure which, according to the requirements for simple structure as now defined, affords to the best solution. In other words, we endeavour to determine the system of hyperplanes which is most overdetermined.

Oblique Structure

Spearman's original factor model included a general factor and one factor specific for each variable. The specific factors were, by definition, mutually uncorrelated and uncorrelated also with the general factor. When investigators became principally interested in finding group factors, the question arose whether they could be assumed to be correlated with one another or not.

If factor analysis is employed only for finding purely descriptive factors suitable for representing a certain correlation matrix, then a model with uncorrelated factors is often easier both, to comprehend and to use. Moreover, uncorrelated factors are, from the viewpoint of mathematical statistics, considerably simpler to handle. To use uncorrelated factors in all cases would however, according to Thurstone (1947, p. 139), be a serious limitation:

> Among statisticians and psychologists there is a rather general belief that if human traits are to be accounted for by any kind of factors, then these factors must be uncorrelated. This belief has its origin in the statistical and mathematical convenience of uncorrelated factors and also in our ignorance of the nature of the underlying structure of mental traits. The reason for using uncorrelated reference traits can be understood, but it cannot be justified. Height and weight are two useful measures of body size, even though they are correlated. If we should use, instead, two linear combinations of height and weight which are uncorrelated in the general population, we might find such measures more convenient in statistical theory, in that the cross products would vanish and we could find a certain satisfaction in that the two measures are independent; but they would be awkward to think about.

If one wishes to find primary factors referring to "functional unities", one has to take into account that these can be correlated with one another. They can be connected with each other through, for example, the influence of heterogeneity of the population, or influence of growth. In a structure of abilities consisting, for example, of a numerical, a verbal, an inductive, and a spatial factor, it is fully possible that they are mutually connected with one another. In any case, the method of analysis must be open to this possibility.

It is entirely reasonable that factors which are uncorrelated or low-correlated in one population will be more highly correlated in another population and vice-versa. If, for example, the structure of abilities in one population is relatively homogeneous with regard to the level of general intelligence, then the group factors of ability in this study could have a considerably lower correlation with one another than in a sample from a population with greater variations in the general level of intelligence. Examples of this are factor analyses of college students compared with those of representative samples from the entire population. If the factor pattern is to be invariant, it is necessary to allow for changes in the correlations between the factors. This has recently been pointed out also by the statisticians Anderson and Rubin (1956, p. 134).

A drawback that has been emphasized, for example, by Reyburn-Raath (1949), is that oblique structure opens up possibilities for still more alternative solutions than only orthogonal structure does.

Whether correlated factors are to be accepted in a certain field is a question that will finally be settled by studies of the factors also in nonfactorial contexts.

The Selection of Test Battery

When the structure in a certain field is to be investigated, all the available knowledge should of course be used. On this basis we may perhaps get an approximate idea of the number of dimensions within the field and of the hypotheses regarding some of the factors. Such hypotheses, on which the selection of tests is based, are always more or less implicit. A basic principle in the selection of tests for the battery is, that the different parts of the field should be covered by tests that are correlated, but at the same time, experimentally independent. A rule which is often applied, is that more than two tests loaded with a certain factor, are needed to locate the hyperplane related to this factor. This is done in order to make the factor overdetermined and easier to interpret.

A major principle in the selection of tests is that none of them shall require all the dimensions that are included in the field. This is how Thurstone (1947, p. 179) expresses it:

> If the test battery is a hodgepodge of complex tests, there may not exist in the configuration any clues as to what the underlying functions may be; but this possibility is dependent on the intuitions of the experimenter in assembling a significant set of measurements rather than on factorial theory and method as such.

The results of the analyses will show, if the test has been selected in such a way that it meets the requirements for simple structure, and so that the solution becomes unique and invariant. If only certain factors become well defined, new hypotheses must be formulated as a basis for selection of new tests which will more clearly define the structure. It is necessary to have a series of factor analyses, of which the first are more explorative, while subsequent analyses deal increasingly with tests built on hypotheses which are based on the results of the earlier analyses.

Another reason why simple structure cannot be obtained may be that the test battery covers too small a field. If, within the field, there is a dominant functional unity that is present in all tests and that cannot be divided into further factors, then the field must be expanded. The test battery must have the property that none of the tests includes all dimensions. Otherwise the principle of simple structure is not applicable to this set of data, and another factor model must be selected.

An often disregarded requirement is that the tests should have high or, in any case, known reliability. Since the factor loadings in a variable decrease in proportion to the square root of the reliability of a variable (Guilford & Michael, 1950), low reliability will result in the factor loadings in this variable becoming so low that they might be overlooked in the rotational procedure or in the interpretation of the factor structure.

There is a mathematical relation between the factor matrix A_c, obtained in the analysis of the correlation matrix corrected for attenuation R_c, and the factor matrix A_m obtained in analysis of the correlation matrix without correction R. This relation follows the matrix algebraic formula

$$A_c = A_m V^{-\frac{1}{2}}, \tag{68}$$

where V is the diagonal matrix which contains reliabilities. This formula, as well as the following ones, are special cases of the general formulae given by Lawley (1940, pp. 69–70).

$$V = \begin{bmatrix} r_{11} & & & & & \\ & r_{22} & & & & \\ & & \cdot & & & \\ & & & \cdot & & \\ & & & & \cdot & \\ & & & & & r_{nn} \end{bmatrix} \tag{69}$$

The relation between R_c and R follows the formula

$$R_c = V^{-\frac{1}{2}} R V^{-\frac{1}{2}} \tag{70}$$

which is the general matrix formula for the well-known attenuation formula

$$r_{\infty\infty} = \frac{r_{12}}{\sqrt{r_{11} r_{22}}}. \tag{71}$$

The above equations imply that one can, theoretically speaking, make a factor analysis of unreliable variables and, after correction for attenuation of the matrix of loadings, obtain the same loadings as an analysis of fully reliable tests should give. This correction for attenuation should be made by dividing the loadings of each unreliable variable by the square root of the reliability of each variable. In practice, the conditions can be more complicated. If, for example, two factors have the loadings of 0.40 and 0.35 in a fully reliable test, those loadings can, because of the lack of reliability of the test, drop to values that would be disregarded in the rotational procedure. This can result in the rotated axes being chosen in a quite different manner from that employed when factors based on fully reliable tests are rotated.

Communality Estimates

Communalities, instead of unities, are included in the diagonal of the correlation matrix, if the analysis is to be limited to the common part of the variances. At present, there is no fully satisfactory method for determining communalities. A mathematically acceptable solution seems to require that the rank of the matrix R is determined in advance (Rosner, 1948). A number of approximative methods have been suggested for estimating the communalities, before the analysis of the correlation matrix starts. These more or less imperfect methods will not be discussed here. They are discussed, for example, by Thurstone (1947, pp. 282–318).

It should also be pointed out that the wider application of electronic computers will make it possible to find suitable communality estimates, through iteration, without entailing too much work. This iteration requires, however, a hypothesis as to the rank of the correlation matrix.

Lately, other methods have been added that do not require communality estimates made in advance. An example of this is Lawley's method No. I. If a certain rank is assumed, communalities are obtained as bi-products of the analysis itself. Lawley's method, however, requires much work, unless electronic computers are available.

Whittle's (1952) methods, which will be discussed in more detail in Chapter VIII, require only approximate knowledge of the proportional size of the communalities in relation to each other. How approximate this must be, is not known yet, since experience with the methods is rather limited.

Thurstone, in cooperation with Bargmann and Henrysson, has tried to use the principal component method for correlation matrices with zeroes in the diagonal instead of communalities (Thurstone, 1954 b, and Bargmann, 1955).

As an introduction to this approach, the following notes on the principal component method for a correlation matrix with communalities in the diagonal may be useful.

In a principal component solution, the loadings must follow the relation

$$a_{jd} = \frac{\sum_{t=1}^{n} a_{td}\,\bar{r}_{jt}}{\sum_{t=1}^{n} a_{td}^2} \quad (j, t = 1, 2, \dots n;\ d = 1, 2, \dots m). \tag{72}$$

In this formula $\bar{r}_{jt}$ is a residual correlation after d–1 factors have been taken out from the matrix R.

At the same time, the principal component loadings follow the relation

$$\sum_{t=1}^{n} a_{td}\,a_{te} = 0 \quad (d, e = 1, 2, \dots m;\ d \neq e). \tag{73}$$

If $\sum_{d=1}^{m} a_{jd}\,a_{td}$ instead of r_{jt} is inserted in the formula 72, the following formula will apply for loadings of the first factor

$$a_{j1} = \frac{\sum_{t=1}^{n} a_{t1} \sum_{d=1}^{m} a_{jd}\,a_{td}}{\sum_{t=1}^{n} a_{t1}^2} = \frac{\sum_{d=1}^{m} a_{jd} \sum_{t=1}^{n} a_{t1}\,a_{td}}{\sum_{t=1}^{n} a_{t1}^2}. \tag{74}$$

This formula can be divided in the following manner

$$a_{j1} = \frac{a_{j1} \sum_{t=1}^{n} a_{t1}^2}{\sum_{t=1}^{n} a_{t1}^2} + \frac{\sum_{d=2}^{m} a_{jd} \sum_{t=1}^{n} a_{t1}\,a_{td}}{\sum_{t=1}^{n} a_{t1}^2}. \tag{75}$$

Since $$\frac{\sum_{t=1}^{n} a_{t1}^2}{\sum_{t=1}^{n} a_{t1}^2} = 1 \quad \text{and} \quad \sum_{t=1}^{n} a_{t1}\,a_{td} = 0$$

the final result is that $a_{j1} = a_{j1}$.

The main idea in Thurstone's method of factor analysis without communalities is, to use a principal component solution so modified that the requirement $j \neq t$ is imposed on the relation

$$a_{jd} = \frac{\sum_{t=1}^{n} a_{td} \quad {}_{jt}}{\sum_{t=1}^{n} a_{td}^2} \qquad (j, t = 1, 2, \dots n).$$

This means that the communalities are zero in the matrix R. Through the addition of this requirement, one of the products will disappear in the product sum $\sum_{t=1}^{n} a_{t1}\, a_{td}$ of formula 75, namely, the one for which $t = j$. This reduced product sum will no longer be equal to zero for the ordinary principal component loadings calculated from the matrix R with communalities. Principal components analysis, with and without communalities, will therefore give the same results only when the rank of the matrix R is equal to one. In this case, the term after the plus sign in formula 75 will disappear and there will only remain

$$a_{j1} = \frac{\sum_{t=1}^{n} a_{j1} \bar{r}_{jt}}{\sum_{t=1}^{n} a_{t1}^2} = a_{j1} \frac{\sum_{t=1}^{n} a_{t1}^2}{\sum_{t=1}^{n} a_{t1}^2} = a_{j1} \qquad (j, t = 1, 2, \dots n;\ j \neq t). \qquad (76)$$

When the rank is higher than one, the differences between the loadings according to the two methods will often be large.

In connection with a study of Lawley's maximum likelihood method I, Henrysson (1950) constructed 12 artificial covariance matrices of nine variables, each containing 200 observations. The observations were based on numbers taken from Wold's Random Normal Deviates (Wold, 1948). The covariance matrices have been converted into correlation matrices in connection with the following study. The expected structure of each variable is very simple, containing one general and one unique factor according to the equation

$$z_{ji} = \frac{\sqrt{2}}{2} F_{1i} + \frac{\sqrt{2}}{2} U_{ji} \qquad (j = 1, 2 \dots 9;\ i = 1, 2 \dots 200). \qquad (77)$$

The correlation matrices are given in the appendix. From each of the 12 correlation matrices, one factor has been taken out according to each of the principal component methods, with and without communalities. The squares of the loadings, found by the method without communalities, have been used as communalities in the method with communalities. The results of the two methods, which are given in the appendix, agree closely. Out of the $12 \times 9 = 108$ pairs of loadings, no loadings in a pair differ by more than 0.001.

Such a good agreement cannot be expected when the rank is higher than one. This is illustrated in a correlation matrix known to be of rank two. Lawley and Swanson (1954, p. 76) have published such a matrix. If the theoretical values of the communalities are used in the diagonal, the correlation matrix will be practically empty after two factors are taken out by the principal component method with communalities. The loadings are to be found in table 5. The residual correlations lie within the limits of rounding errors, since none of them has a value over 0.001.

TABLE 5. *Matrices of factor loadings after principal component analysis with and without communalities*

Variable	Loadings after analysis with communalities		h^2	Loadings after analysis without communalities		h^2
	I	II		I	II	
1	.741	.560	.863	.693	.565	.799
2	.869	.351	.878	.861	.376	.883
3	.664	.670	.890	.600	.660	.796
4	.894	.003	.799	.933	.017	.871
5	.818	−.405	.833	.806	−.391	.803
6	.813	−.483	.894	.785	−.459	.827
7	.748	−.593	.911	.700	−.550	.793
			6.068			5.772

In the analysis without communalities of the same matrix, the loadings are quite different, as is shown in table 5. The residual correlations, after two factors are taken out, are larger in this case, as can be seen in table 6. These results illustrate what has been shown algebraically, namely, that the two methods give the same result only when the rank is one in the correlation matrix R.

Table 6. *Residual matrix after two factors by principal component analysis without communalities*

$$\begin{bmatrix}
 & .032 & .078 & .007 & .041 & .046 & .048 \\
.032 & & .047 & -.030 & .022 & .034 & .045 \\
.078 & .047 & & .024 & .046 & .047 & .041 \\
.007 & -.030 & .024 & & -.015 & .002 & .023 \\
.041 & .022 & .046 & -.015 & & .049 & .073 \\
.046 & .034 & .047 & .002 & .049 & & .093 \\
.048 & .045 & .041 & .023 & .073 & .093 &
\end{bmatrix}$$

In conclusion, the impression gained is that the principal component method without communalities is not very useful except for rank one. In this case, other simple methods are available, for example, Spearman's, which does not require knowledge of the communalities.

The only method, which is useful when the rank is over one, and which does not require communality estimates made in advance or any assumptions about their size, is, at present, Lawley's method I, which will be discussed in Chapter VIII.

Analytical Methods for Rotation to Simple Structure

The analysis of a correlation matrix into a simple structure factor matrix must, at present, be made in two steps. First, a preliminary factor matrix is determined, if possible by means of an efficient method of analysis. This method should include a test of significance so that the order of the factor matrix can be kept down. This very important problem will be considered separately in Chapter VIII.

The next step is the rotation, which is to transform the preliminary factor matrix into a new matrix which, as far as possible, meets the requirements for simple structure. Thurstone and his followers have generally made this transformation by graphical methods. These procedures include subjective elements that require experience in the interpretation of graphical representations, and are also very tedious. (Thurstone, 1947, pp. 194–410.)

The many applications of factor analysis to correlation matrices of high rank and the production of highly effective calculators have made the need for objective rotation methods still more essential. Even if the five requirements for simple structure would be sufficient for producing overdetermined simple structure solutions, they are, however, subjective and ambiguous and are not summed up in a single index. Therefore, many attempts have been made to find analytical methods that give a mathematically unique definition of simple structure. At present, this is a field of factor analytical theory, which is undergoing rapid development.

A group of analytical methods is based on "the quartimax criterion" (Saunders, 1953; Pinzka-Saunders, 1954; Carroll, 1953; Neyhaus-Wrigley, 1954; and Ferguson, 1954). These methods, generally speaking, are based on the idea of maximizing the sum of the fourth power of the rotated factor loadings. Pinzka-Saunders (1954), for example, achieve this, by the function

$$K = \sum_j \sum_k a_{jk}^4 / (\sum_j \sum_k a_{jk}^2)^2 \qquad (j = 1, 2 \ldots n;\ k = 1, 2 \ldots m), \qquad (78)$$

where a_{jk} is a loading in a matrix of oblique factors. A criterion of this kind tends to maximize the number of high and low loadings by avoiding those of medium size (Saunders, 1953, p. 5). Different improvements of the above mentioned criteria are under way, and tests are going on with them by use of empirical data. At present it is too early to decide if some of these criteria are fully satisfactory, at least for solving certain types of problems. Experience indicates, however, that in certain cases criteria of this kind tend to give rise to general factors also when rotation to simple structure without a general factor is possible. (Neyhaus-Wrigley, 1954, p. 90.)

Thurstone (1954a) has also tried to develop an analytical method which is largely independent of subjective judgements in the rotational work. This method is an extension of his single-plane method of rotation (Thurstone 1947, pp. 216–224). When the single-plane method is used, certain tests are selected for inclusion in the hyperplane at a relatively late stage of the rotation. By the introduction of the new analytical method, this procedure has been made more effective so that it can start directly from the preliminary solution, for exampel, from a centroid solution.

The main principle is that the tests are given weights, which are higher, the closer the tests are to the expected hyperplane. Thurstone's analytical method has been improved by Bargmann (1953). He has found that the method often gives results that must be improved by further rotation (Bargmann, 1953, p. 4). This is also the author's experience of the method.

Another attempt to solve the rotation problem has been made by Tucker, who starts out from the following definition of simple structure (Tucker, 1955, p. 218):

> A simple structure is constituted by the hyperplanes for a set of r linear constellations of dimensionality $(r-1)$.

Tucker's method is based on the idea of maximizing the number of vectors, located within a marginal space of some defined and limited width, from each one of the hyperplanes of dimensionality $(r-1)$. It is too early yet to express an opinion on the effectiveness of Tucker's new method. He also emphasizes that the test battery must be very well-designed if his method is to work properly.

In conclusion: The impression of the various analytical methods for rotation to simple structure is that, at present, none of the methods results in as good a simple structure as the graphical rotation methods give, if these are used skillfully by an experienced investigator. While waiting for improvements of the analytical rotation methods, these should be used for finding preliminary factor matrices, which then ought to be improved by graphical rotation.

CHAPTER VII

THE PROBLEM OF INVARIANCE

The Concept of Invariance

It was stated earlier in this monograph that invariance is an important requirement on explanatory factors. Even if one factorial analysis were to afford a unique simple structure solution, this cannot be taken as sufficient proof that the factors are primary in the sense that they have explanatory properties. Though the design and the test selection are carefully made, some of the factors may still be artificial products of the very constellation of tests contained in the battery. A single factor analysis yields more or less unverified hypotheses as to factors which must be proved invariant through other factor studies employing other tests and in respect of other populations, before it can be said that the factors found have the generality required of factors with explanatory properties.

Comparisons between factor matrices found in regard to different populations presuppose that the factor pattern is similar, in the sense that approximately the same common factors are present in the different populations. Even if the same test is given to different populations, it does not mean that the same factors are involved in solving this test. Tests which, for a population of ten-year-old children, for example, measure, in the majority of cases, factors of intelligence may, in the case of adults, measure perceptual speed. Consequently populations that are to be compared should not be too different, for example, in age, as otherwise different factors may be present.

Thurstone has discussed the requirements for invariance, as summarized above on page 46. Four cases of invariance were mentioned. In connection with the following discussion of methods of studying invariance, it may be useful to distinguish between six cases, by dividing each of the two cases of changed battery, into cases of partly and of entirely changed battery. The six cases refer to when the test battery is the same, or partly changed, or entirely changed, and when the populations are either the same or different.

This is evident from the following table.

TABLE 7.

	Populations	
Test battery	Same	Different
Same	(1) numerical invariance	(2) configurational invariance
Partly changed	(3) numerical invariance	(4) configurational invariance
Entirely changed	(5) configurational invariance	(6) configurational invariance

Each of the six cases gives rise to different problems:

CASE 1.—When the same test battery is used for the same population, there should be numerical invariance in all samplings of the population. That means that the factor structure found in the different samples should be identical within the limits of the sampling errors. Studies of this kind of invariance require a knowledge of confidence limits for differences between factor loadings calculated for separate samples from the same population.

CASE 2.—If the same test battery is used for different populations, then configurational invariance may be expected to take place. This signifies that the change of population affects the size of the factor loadings, but in proportion to the changes in the variance of the different tests administered to the population. This implies that if a test has zero-loading in a certain factor in one population, it will have zero-loading in the same factor also in another population. This is a necessary condition if the same test vectors are to define the same hyperplane. The configuration of hyperplanes must, in other words, be the same even if the angles between the normals, which define them, are changed. This can be illustrated by the extended vector method (Thurstone, 1947, pp. 225–258), in cases where there are three common factors. If simple structure exists, the heads of the test vectors should coincide with the sides of a triangle, for example, *ABC* in Figure 5.

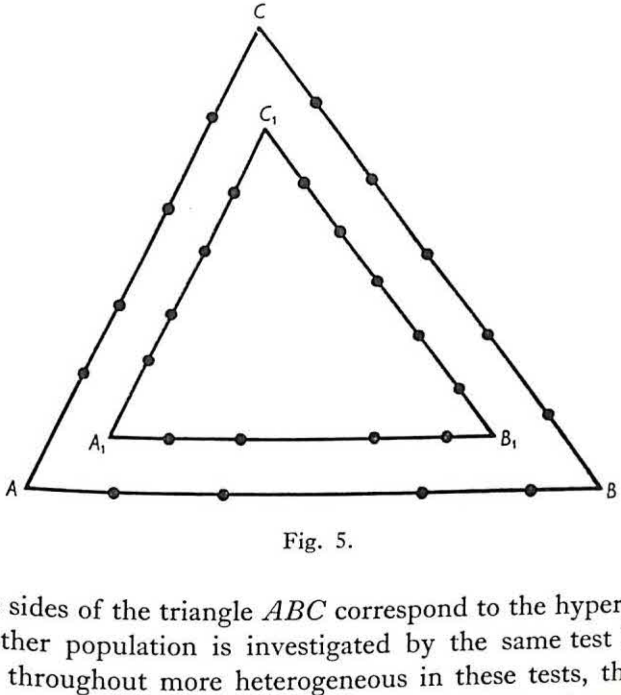

Fig. 5.

The sides of the triangle *ABC* correspond to the hyperplanes. If another population is investigated by the same test battery and is throughout more heterogeneous in these tests, then the tests have higher intercorrelations. This will under the assumption of configurational invariance result in the test vectors coming closer to each other and consequently forming a smaller triangle, for example, $A_1B_1C_1$.

Thurstone has concentrated his attention on determining to what extent approximate zero loadings occur in the same places in two factor matrices. This is too subjective a method and does not make use of all the information obtainable from the factor matrices.

Stricter methods of studying invariance have been suggested, among others, by Rasch (1953). He points out that the lack of numerical agreement between factor matrices is due to the fact that the normalizing of test scores and of factor scores has been made separately for each population. For, the result of using correlations is that all tests and factors included in the test battery have variances equal to unity. If the loadings in the different factor matrices are to be numerically the same for a certain test, then test variances and factor variances must be proportional to changes between one population and another in respect of the actual test variances. Factor analyses must, in such a case, be made of covariance matrices, in which variances and covariances of a certain test are calculated by the same unit of measurement, irrespective of the population for which the test is used. The factor loadings thus determined are expected to be of the same size irrespective of the population, and may only vary within the limits of sampling errors, unless a factual difference exists between the factor structure of the populations. The statistical testing of the differences will be more complicated than in case 1, since the populations are different from each other.

Rasch also suggests a method for the simultaneous analysis of a number of such covariance matrices, but he has not given any solution which is useful from a practical point of view.

CASE 3.—If the test batteries, used on samples from the same population, include tests of which only some are the same, the numerical invariance requires that each of the tests that is common to two or more batteries retains its factor pattern. In other words, the hyperplanes must have the same position, even if other test vectors contribute in determining their position. If the hyperplanes are the same, this indicates that the factors are not merely accidental products of the particular configuration of tests in a certain battery. The problem of sampling errors is similar to that in case one, but must be concentrated on the tests that are common to the batteries.

CASE 4.—Here partly different test batteries are used on different populations. Case 4 is a combination of cases 2 and 3 above, since both the test battery and the population are changed. The configurational invariance signifies, in this case, that the same hyperplanes are to be localized, but partly with the help of other test vectors; at the same time also the angles between the hyperplanes can be changed.

It should be possible to carry out a statistical examination of differences in factor structure between factor matrices based on test batteries which are partly different, i.e. in accordance with cases 3 and 4, by employing those tests which are common to the batteries. The procedure ought then to be similar to that used for cases 1 and 2 respectively, but is likely to be less certain since, as a rule, a smaller number of tests are included in the examination.

CASES 5 AND 6.—The most convincing form of invariance should, in principle, be obtained through the analysis of two completely different test batteries. The risk that an accidental constellation of tests might give rise to artificial factors would then be minimal. When one test battery is compared with another which is composed of completely different tests there are, however, no mutually corresponding loadings in the two factor matrices. A comparison must therefore be based either on a subjective impression that a certain factor is the same in both cases or on some form of indirect comparison. The latter should, in principle, be possible through certain other tests that are correlated with at least some tests from each of the batteries.

Methods for Finding Invariance

A number of attempts have been made to devise methods for testing invariance. Some of these methods are mainly limited to investigations of configurational invariance. Certain of these methods are based on the fact that the rows of loadings of a factor, which is expected to be the same in two different factor matrices, should have high correlation with each other. Other methods compare, for example, by means of probability models, the profiles of the factor supposed to be invariant in the two factor matrices (Cattell–Baggaley, 1954).

More exact tests of invariance must make full use of the numerical values of the loadings. For such tests of numerical invariance, a common unit of measurement for a test is required, even when the test is given to several populations. The factor analyses then have to be based on covariances instead of correlations. The use of correlations means that each test, irrespective of its absolute variance, is converted into scales with the variance of unity in all test batteries. If studies of two populations that have different variances in one or more tests, are to be compared, the correlation matrices ought to be replaced by covariance matrices, where the covariances should adopt values that correspond to the relation between the absolute variances of the test in respect of the two populations. If, for example, a certain test has the standard deviation of 20 points in the first population and 10 in the second, then the correlations that include the test should be multiplied by the weights which are twice as big for the first population as those for the second. Such a change in the standardizing of tests can be made also at a later stage of the analysis, viz., in the matrices of factor loadings calculated from the correlation matrices. This means that the loadings in the test are multiplied by their respective weights.

This type of criterion of numerical invariance between factor solutions in two different populations has been developed by Ahmavaara (1954b).

He takes as his starting point the discussion between Thurstone and Thomson on the influence of multivariate selection on invariance. Thurstone has given examples showing that if oblique factors are accepted and simple structure is used, then the factor pattern can be invariant after changes of the variances have taken place in some variables, which explains the ensuing changes in the intercorrelations.

Ahmavaara (1954a) shows, by employing the Pearson-Aitken's selection formulae, that a simple structure solution, according to the conditions stated by Thurstone for multivariant selection, is invariant. The loadings are changed in proportion to the changes in the standard deviations of the tests and zero loadings are consequently unchanged. He also gives a method, based on the formulae for correction after selection, for comparing factor matrices referring to two different populations with each other. (Ahmavaara, 1954b.) First the sizes of variances and covariances in one of the matrices are adjusted so that a variable common for the two batteries is measured by the same units of measurement. Ahmavaara is looking for a comparison matrix that gives the best numerical invariance between the two matrices of factor loadings. The comparison matrix is of the same order as the number of factors and the elements of this matrix express the factors of the first study as linear combinations of the factors in the other study. If the comparison matrix is a unit matrix then the invariance is complete. Ahmavaara has shown how his method works in several examples. He has, for instance, compared the results obtained in L. L. Thurstone's study *Primary Mental Abilities* with those in L. L. and Th. Thurstone's investigation *Factorial Studies of Intelligence*. Ahmavaara found that the invariance is good only for some of the factors. Ahmavaara (1957) has recently published several other applications of the method.

Ahmavaara has given the problem of invariance investigations a stricter formulation than was formerly the case. His approach is, however, mathematical rather than statistical and he has not considered the problem of sampling errors when comparing factor studies made on two different samples (Ahmavaara, 1954b, p. 43).

Ahmavaara's method is applicable when either the same battery, or a battery which is partly the same, is used for samples from the same population or from different populations. These cases correspond to cases 1–4 in Table 7.

Tucker (1956) has worked out a method for comparing factor matrices which are based on two completely different test batteries given to the same sample. This corresponds, in principle, to case 5 in Table 7. The two test batteries should be expected to measure the same common factors, if the method is to work properly. The correlation matrices for the two batteries 1 and 2, are designated by R_{11} and R_{22} respectively. In addition, there is a matrix $R_{12} = R_{21}$, which includes the correlations between the tests in battery 1 and the tests in battery 2. The correlation matrix that includes all intercorrelations will then be as follows

$$R = \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix} \tag{79}$$

Matrix R_{12} is factor analyzed by a specially adjusted principal factor solution. No communality estimates are needed since communalities are contained only in R_{11} and R_{22}, which are not included in this part of the analysis. Tucker has worked out a test of significance for the residuals in R_{12}, but he himself is doubtful regarding its usefulness.

The preliminary factor analysis of R_{12} results in a factor matrix, which can be divided into two parts: A_1 corresponding to battery 1, and A_2 corresponding to battery 2. The factors found in both A_1 and A_2 are invariant in the two batteries. Each of the matrices A_1 and A_2 can then be rotated separately to simple structure, and the main interest in this type of invariance analysis is to study the agreement between the corresponding factors in the two rotated matrices. This is achieved by a serie of correlations between corresponding factors in the two rotated solutions. Tucker calls these correlations "factor reliability coefficients", but until further investigations have been carried out, he is unable to express an opinion as to what values of this coefficient are to be considered as high or low. The coefficient values tend to be low if the reliabilities and communalities of the separate tests are low.

This method has many advantages, for instance, it does not require communality estimates, it employs an efficient method of analysis, and it contains measures of invariance. One disadvantage is, that it does not make full use of the information contained in R_{11} and R_{22}, as does an analysis of the whole matrix R, according to a conventional method. Therefore, the method is more useful when R_{12} actually includes big loadings in the common factors, which are found in R_{11} and R_{22} if they are analyzed separately.

Simultaneous Rotation to Invariant Solutions

Cattell has considered the invariance problem from another angle. He is concerned with methods for finding one simple structure solution, which at the same time, can be applied to two or more correlation matrices. He describes this "so-called principle of parallel proportional profiles" in the following way (Cattell & Cattell, 1955, p. 84):

> The basic assumption is that, *if a factor corresponds to some real organic unity, then from one study to another, it will retain its pattern, simultaneously raising or lowering all its loadings according to the magnitude of the role of that factor under the different experimental conditions of the second study*. No inorganic factor, a mere mathematical abstraction, would behave in this way, within the realm of orthogonal factors. The principle suggests that every factor analytic investigation should be carried out on at least two samples, under conditions differing in the *extent* to which the same psychological factors (working as independant, orthogonal influences) might be expected to be involved. We could then anticipate finding the 'true' factors by locating the unique rotational position (simultaneously in both studies) in which each factor in the first study is found to have loadings which are proportional to (or are some simple function of) those in the second: that is to say, a position should be discoverable in which the factor in the second study will have a pattern which is the same as in the first, but is stepped up or down.

Cattell has devised an analytical method for simultaneous rotation to simple structure (Cattell & Cattell, 1955). This method has, however, some drawbacks, which the authors also have pointed out. Thus, it is only applicable for rotation to orthogonal solution. The basic principle of simultaneous rotation to simple structure seems both important and promising, but e.g. because of the last-mentioned reason, we can expect that there will be improvements in this method.

Tucker (1951) has also dealt with the problem of carrying out simultaneous rotation to invariant solution. He has chiefly been interested in devising a method for making the preliminary factor matrices what he calls "congruent with each other". This signifies that he tries to rotate the matrices so that each of the factors will, if possible, have the same content. Unlike Cattell, he does not include restrictions, such as simple structure, in the matrices. Tucker considers that these matrices can, if reasonably congruent, be rotated together to a significant solution, e.g. according to the principle of simple structure.

The starting-point for Tucker's method is, to use his symbols, the factor matrices F_{jmA} and F_{jMB}, which are two preliminary and unrotated matrices with factor loadings, for example, two centroid matrices. The subscript j is used to designate the tests with unit standard deviations in the separate studies, and m and M respectively designate the reference vectors in the two investigations. A and B are used to designate the two studies.

First it is necessary to determine the series of weights by which the different tests are multiplied, in order to express each separate test in the same unit of measurement in the two studies. These weights are included in the two diagonal-matrices D_{JA} and D_{JB}. The subscript J designates the tests after adjustment of the standard deviations. The next step comprises the matrix multiplications

$$\begin{aligned} F_{JmA} &= D_{JA} F_{jmA} \\ F_{JMB} &= D_{JB} F_{jMB} \end{aligned} \tag{80}$$

By this operation the two factor matrices F_{JmA} and F_{JMB} are made comparable in respect of units of measurement.

Tucker then uses the transformation matrices T_{mrA} and T_{MrB} which are of such a nature that when

$$F_{jrA} = F_{JmA} T_{mrA} \quad \text{and} \quad F_{jrB} = F_{JMB} T_{MrB}, \tag{81}$$

the factor matrices F_{jrA} and F_{jrB} are congruent. Subscript r designates factors in the new solution. The differences between these matrices ought to be negligible for tests which overlap in the two studies, provided there is invariance.

Tucker tests this by means of "a coefficient of congruence for each factor":

$$\phi_r = \frac{\sum_J f_{JrA} f_{JrB}}{\sqrt{(\sum_J f^2_{JrA})(\sum_J f^2_{JrB})}}. \tag{82}$$

The symbols f_{JrA} and f_{JrB} refer to loadings in factor r in the matrices F_{JrA} and F_{JrB}. ϕ_r will be unity when the congruence is complete. The coefficient lacks a test of significance, and Tucker

(1951, p. 43) does not wish to state any acceptable minimal value of the coefficient of congruence, until adequate experience in the application of the method has been gained.

Tucker points out that mathematical congruence is a necessary, but not a sufficient, requirement for identity of the factors between two studies. This identity becomes more reliable, the larger the number of factors common to the two studies. A psychological analysis of the content of the test should also be made, in order to indicate that the factors are not only mathematically but also psychologically congruent.

When the congruent factor matrices have been determined, they can then be rotated together, for example, according to the principle of simple structure, to two new congruent factor matrices. These are then invariant in relation to each other.

Conclusions about the Methods of Finding Invariance

Differentiation between six cases of invariance has been made; cases which can more or less easily be established on a common basis. To rely on subjective impressions that a certain factor is the same in several studies is an unsatisfactory way of investigating invariance. A more objective method is to investigate whether configurational invariance exists. Still more objective and efficient are the methods which investigate the degree of numerical invariance after each test has been adjusted to the same unit of measurement in all studies in which it is included.

Ahmavaara, Tucker, and Cattell have all considered this type of problem and have contributed towards its solution. At present there is no completely satisfactory method which, after the proper adaption of units of measurement in the tests, can show invariance between different factor solutions, and which at the same time is provided with an adequate test of significance. The methods mentioned above, can no doubt be improved and still further developed in these and other respects.

A possible solution would be to design the investigations of invariance as tests of hypotheses, e.g. as follows. On the basis of the available knowledge regarding both the tests to be analyzed and regarding the earlier factor analyses, hypotheses are formulated as to the common factor structure and, if possible, also as to the numerical values of the loadings. Each test is standardized with regard to the same unit of measurement in all the factor studies. This implies comparisons between factor matrices based upon analyses of co-variance matrices. The hypotheses are tried out on different test batteries and populations. The hypotheses should then step by step be more exactly formulated and improved. The problem of the test of significance for such a procedure is mentioned in the last section of Chapter VIII.

Non-factorial Validation

Numerical invariance is both an important and necessary condition for giving explanatory significance to the factors found but it is not entirely sufficient by itself. A requirement stated by many investigators is that the factors should also be supported by investigations other than factor analytical studies. Thurstone (1947, p. 334) formulates this as follows:

> The acceptance of primary factors in psychological investigation will depend on the extent to which they are fruitful in non-factorial psychological research.

This implies that we should be able to use the factors, in the formulation of testable hypotheses as to relations and regularities, in as wide a context as possible.

If, for example, a certain disease, injury or operative incision, leads to the loss or weakening of a certain factor, but without affecting other factors, it indicates that a functional unity has been found. Growth studies, showing that the development of a certain factor is separate from that of other factors, would be another type of non-factorial confirmation.

Unfortunately, there are not many such investigations available. Thurstone, for example, has made certain attempts along these lines in his unpublished *Growth Studies of Primary Mental Abilities*.

Eysench (1954, pp. 107–142) has tried to obtain confirmation of political-attitude factors by comparing them with political behaviour. He made a factor analysis of correlations between questions referring to political attitudes, and found two factors. Besides this, he calculated the differences between Conservative and Socialist voters, in percentage of positive answers to the same questions. He found a fairly high correlation between the loadings of the first factor and the differences in percentage of positive answers. This he regarded as evidence that the first factor could get a meaningful interpretation as a factor of Radicalism.

A number of studies of colour vision have been made by employing factor analytical methods. As an example of this we may mention Ekman's (1954) study, in which paired comparisons between colours referring to light of certain wave lengths were studied by an application of factor analysis, called *Similarity Analysis*. Ekman found five colour factors which covered, to a large extent, the wave-lengths which Granit, by means of psysiological techniques, found to be connected with different receptor groups of colour vision.

The impression is, however, that on the whole very little has been done towards systematic non-factorial validation of factor analytical results.

CHAPTER VIII

FACTOR ANALYSIS FROM A STATISTICAL POINT OF VIEW

The Need of Statistically Acceptable Methods

Factor analysis has been developed mainly by psychologists as a mathematical technique for the treatment of test results and other measures of individual differences. Statisticians have worked out solutions for only some of the statistical problems involved, but are now concerning themselves more and more with the whole field. The tendency has often been for psychologists themselves to develop factorial methods, which are inefficient from a statistical viewpoint, for converting a correlation matrix, with the exception of negligible residual correlations into a matrix of factor loadings with low rank. It is only subsequently that they have tried to deduce the sampling distributions, and to determine confidence intervals for the residual correlations, in order to make it possible to test whether these really are negligible, and to decide when to stop factoring the correlation matrix. Because the psychologists have not made full use of current statistical theory, the confidence limits for the residual correlations are very often given an unsatisfactory basis. Many of these attempts have been characterized by their authors as approximations only to be used for want of anything better. Burt (1952), among others, has discussed these methods thoroughly.

These conditions have led many specialists in statistical theory to be both hesitative and critical in respect of factor analysis and have also caused them to endeavour to persuade the psychologists to formulate their problems differently (Hotelling, 1942).

Some statisticians have only recently accepted the distinction in psychological research between component and factor analysis, which is made in Chapter I. One reason for this is probably that for a long time only one method existed, which is satisfactory from a statistical viewpoint, and which is useful in component analysis, viz. the principal component solution. Any counterpart to this, which is useful in factor analysis, has been lacking for a very long time. Efforts have been made, however, to use principal component or related methods also for problems of a factor-analytical nature.

Before discussing these approaches, it might be well to repeat the distinction between component analysis and factor analysis, which was already made in Chapter I.

In component analysis two cases may be distinguished.

$$(1\,\mathrm{a}) \quad z_{ji} = \sum_{k=1}^{q} a_{jk} F_{ki} \quad (j=1, 2 \ldots n;\ i=1, 2 \cdots N;\ k=1, 2 \ldots q). \tag{83}$$

Each test score is here considered as a weighted sum of q components. These in their turn are transformations of the n tests, and make possible the exact representation of the test scores. The component loadings are found in matrix A, which follows the relation $R_1 = A \Phi A'$. In psychological applications, q will probably be equal throughout to the n tests of the battery, since each test will, in practice, partly measure something specific.

$$(1\,\mathrm{b}) \qquad z_{ji} = \sum_{k=1}^{q} a_{jk} F_{ki} + a_{j\varepsilon}\, \varepsilon_{jk}. \tag{84}$$

In this case, each test score is impaired by errors of measurement. The component loadings form a matrix A_2. As in case 1a, q in this matrix is often equal to n. The n loadings for the error factors constitute the diagonal matrix A_ε. In case 1b the following matrix relation is valid.

$$R_1 = A_2 \Phi A_2' + A_\varepsilon^2. \tag{85}$$

Also in factor analysis two cases can be distinguished. They are called 2a and 2b.

$$(2\,\mathrm{a}) \qquad z_{ji} = \sum_{k=1}^{m} a_{jk} F_{ki} + a_{js} S_{ji}. \tag{86}$$

This model corresponds to case 1a in component analysis but also includes one specific factor in each test, the loadings of which are found in the matrix A_S. The different common factor loadings a_{jk} correspond to common factors in the matrix A_m. The introduction of specific factors will, as experience shows, make it possible, as a rule, to reduce the number of common factors m to considerably below the number of tests n.

The relation between the correlation matrix R_1 and the factor loading matrices is

$$R_1 = A_m \phi A_m' + A_S^2. \tag{87}$$

In addition

$$R = A_m \phi A_m'. \tag{88}$$

The matrix R is a correlation matrix with communalities in the diagonal.

(2b)

$$z_{ji} = \sum_{k=1}^{m} a_{jk} F_{ki} + a_{jS} S_{ji} + a_{j\varepsilon} \varepsilon_{ji} \tag{89}$$

In this case the test scores also contain errors of measurement, the loadings of which form the matrix A_ε. Then the relation is given by

$$R_1 = A_m \phi A_m' + A_S^2 + A_\varepsilon^2. \tag{90}$$

In addition

$$R = A_m \phi A_m'. \tag{91}$$

The inclusion of specific and error factors in the model, which together are called unique factors, makes it possible to reduce the number of common factors m in the matrix A_m so that they are considerably less than the number of tests n.

The main distinction between factor analysis and component analysis is that the former includes specific factors. In psychological applications this facilitates the reduction of the number of common factors, i.e. the rank of matrix A_m and also makes it easier to obtain invariant solutions.

The distinction between component analysis and factor analysis is conditioned by the nature of the problems in behavioral science. Component analysis as applicable to case 1b, where an error term is included in the model, and both the cases of factor analysis can however, be treated by the same statistical method (Whittle, 1952). This will be shown below.

The ideal solution of the factor analytical problem, from a statistical viewpoint, would be to have an efficient method which, without rotation, leads directly to a psychologically significant solution, i.e. according to the principle of simple structure, and, which, at the same time, permits the development of fully acceptable confidence limits for the residuals and the different loadings.

Maximum Likelihood Solutions and Least Squares Solutions

In their efforts to develop acceptable factorial methods statisticians have used approaches based either on maximum likelihood- or on principal component solutions. In 1940 Lawley published a maximum likelihood method of factor analysis including a test of significance for residual correlations. The factorial method called Lawley's method I, is based upon the assumption that all test scores and factor scores are statistical by nature and have normal distribution. The test of significance is useful for large samples, and by means of Bartlett's (1950) adjustment it can also be used for medium-sized samples.

The method has as its starting point a hypothesis as to the number of common factors, i.e. the rank of the matrix R. On the basis of this hypothesis an orthogonal factor matrix is estimated, corresponding to the population correlation matrix for which the likelihood of occurrence of the observed correlation matrix in the course of random sampling is a maximum. With the aid of the test of significance it is determined whether the residual matrix, when the hypothesized number of factors are abstracted, can be considered to be dependent on sampling fluctuations or not. If not, then the analysis is repeated on the basis of another hypothesis as to rank. Estimates of communalities need not be made in the analysis, but will be obtained as an end product.

The method has been tried out on artificial samples from populations whose factor structure is known and where the "test scores" are normally distributed. Henrysson (1950) used a system of 9 variables containing one common factor, and 12 samples each comprising 200 observations were taken. Lawley and Swanson (1954) used a system of 7 variables containing two common factors and took 8 samples each consisting of 50 observations. They also analyzed the total sample of 400 observations. The results of these studies seem to indicate that both the method and the test of significance work well even for samples of only 50 observations.

Lawley's method I has certain drawbacks. It is built on the assumption of random distribution of factor scores and test scores. Andersson and Rubin (1956, pp. 130 and 145) have shown, however, that for large samples the method can also be used when this assumption is not fulfilled; to what extent it must be satisfied is not yet known.

Whittle (1952) has developed solutions using least squares methods and has shown that this leads to principal component solutions. By this approach he avoids Lawley's assumption, and regards both loadings and factor values as parameters which have to be estimated. Young (1941) has dealt with the same problem in the simple case of rank one. Whittle discusses his method mainly in connection with the model

$$z_{jk} = \sum_{k=1}^{q} a_{jk} F_{ki} + a_{j\varepsilon} \varepsilon_{jk}. \tag{92}$$

This corresponds to case 1b of component analysis. In the article in which he discusses his method, he treats a model which only includes, besides the "common factors", what he calls "residuals" or "accuracies" (Whittle, 1952, pp. 223–225).

Whittle, however, points out that his special form of principal component solution is applicable even to the factor analytical models which include specific factors, i.e.

$$z_{ji} = \sum_{k=1}^{m} a_{jk} F_{ki} + a_{jS} S_{ji} \tag{93}$$

or

$$z_{ji} = \sum_{k=1}^{m} a_{jk} F_{ki} + a_{jS} S_{ji} + a_{j\varepsilon} \varepsilon_{ji}. \tag{94}$$

In the first of these formulas what Whittle calls "residual errors" are represented by $a_{jS} S_{ji}$ and in the second formula by $a_{jS} S_{ji} + a_{j\varepsilon} \varepsilon_{ji}$.

Another way of expressing this is to say that Whittle's approach is valid for the model

$$z_{ji} = \sum_{k=1}^{m} a_{jk} F_{ki} + a_{jU} U_{ji}. \tag{95}$$

This model includes a unique factor, which can be composed in three ways: it may consist of only an error factor, or of only a specific factor or of both an error and a specific factor. The last two cases correspond to factor analytical models, to cases 2a and 2b mentioned above. In Whittle's approach all these three cases are comprised in one statistical method.

When specific factors are included in the model the number of factor loadings will be very large and the problem of estimation will be more difficult (Holzinger-Harman, 1941, p. 265). This view was also expressed in the discussions in connection with the factor analytical symposium in Uppsala in 1953 (*Uppsala Symposium*, 1953, p. 73):

> Dr. Whittle asked if the introduction of a specific factor was necessary, and whether it did not imply a surrender of the essential factor hypothesis. Would a battery of 2000 tests really need 2000 or more factors for its explanation? It is possible that the apparent occurrence of specifics is due to the fact that the number of real factors is large in comparison with the usual size of a test battery.

That Whittle's method is useful also when specific factors are included in the model is pointed out even by Wold (1953, p. 63) who nevertheless also mentions the difficulties incurred by the inclusions of specifics in the model:

> Such specific factors do not enter Whittle's theory, but it is clear that if we wish to do so we can interpret his residual as a specific factor (or alternatively as the sum of a specific factor and a measurement error). This change in the interpretation does not affect the formal development of Whittle's method. It is quite another question, however, whether the introduction of specific factors is at all logical or appropriate from the viewpoint of the fundamental idea of factor analysis, that of explaining a set of test scores in terms of a smaller number of psychological factors.

By employing Whittle's methods one can estimate both factor loadings and factor scores. The only general assumption required is the ordinary one, i.e. that all factors, including even the unique ones are uncorrelated with each other. Whittle's object is to solve A_m in the equation

$$R_1 = A_m A_m' + A_U^2 \tag{96}$$

but he is unable to do so for the general case. However, he considers two special cases of interest, viz. the following (Whittle, 1952, p. 229):

> We can distinguish, however, at least two special cases which deserve further attention. The first is that in which the relative accuracies of the different measurements are known. The measurements may thus be standardised so that all residuals have the same variance, and we are back in the case of the previous section (except that we have now allowed for variation in mean value). The second case is that in which the error variance of a measurement may be assumed proportional to its population variance...."

In the first case, called "the additive case", the assumption is made that all unique variances are alike, i.e.

$$a_{jU}^2 = \text{const.} \qquad (j = 1, 2, \ldots n). \tag{97}$$

Since in most actual cases the unique proportions of the test variances are different, each test variance must be standardized so that all unique variances become equal. This presupposes that instead of the correlation matrix a covariance matrix is considered such as the following

$$R_U^{-\frac{1}{2}} R_1 R_U^{-\frac{1}{2}}. \tag{98}$$

This relation gives the equation system

$$|R_U^{-\frac{1}{2}} R_1 R_U^{-\frac{1}{2}} - \lambda I| = 0. \tag{99}$$

The latent vectors according to this equation system will, after adjustment, correspond to factor loadings.

This solution assumes that the test variances are standardized in such a way that the unique variances are equal. Their size need not be known, however. It is sufficient to know their relative size in relation to each other, so that they can be made equal by changing the sizes of the test variances.

The significance of this is apparent from the coordinate system in figure 6, where the x-axis corresponds to the total test variance, suitably standardized, and the y-axis corresponds to the unique part of the test variance. In the additive case the unique variances should be equal in size regardless of the size of the test variances. Consequently, the unique variances should form a straight line at the same distance from the x-axis. According to Whittle it is sufficient to assume that the line is horizontal. The extent of the distance α from the x-axis does not need to be known.

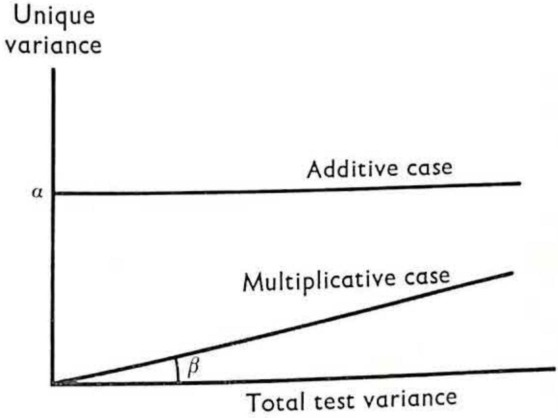

Fig. 6.

In the latter case, called "the multiplicative case", it is assumed that all unique variances are in the same proportion to the respective test variances, i.e.

$$a_{jU}^2 = \text{const.}\ s_j^2 \quad (j = 1, 2, \dots n). \tag{100}$$

This is also illustrated in Fig. 6. The unique variances are changed in direct proportion to the change in the test variances. The relation corresponds to the tangent of the angle β, the size of which does not need to be known. According to Whittle, it is sufficient to assume that the line passes through the origin. The assumption of constant relation means, in this case, that unique variances should be of equal size if a correlation matrix is analyzed, because in such a matrix, all test variances are standardized to the same size. The size of the unique variances need not, as previously mentioned, be known. It is sufficient to assume that the proportion of the unique variance is similar for all tests. The equation system in the multiplicative case will be

$$|R_1 - \lambda I| = 0. \tag{101}$$

The assumption in the multiplicative case, that all unique variances are proportional to the respective total variances, will very seldom be satisfied in psychological and similar applications. The additive case will therefore be of greater interest, at least as regards the behavioural sciences.

Wold (1953) has tried both the additive and the multiplicative cases on artificial samples, when the scores of the two common factors were sampled from the positive half of a normal distribution, and the residuals were based on independent random samples from a normal distribution. The common factors were taken from a truncated distribution in order to make it possible to investigate whether the method works even when the distribution of the common factors diverges considerably from the normal distribution. There were five "tests" and the size of the sample was 50. Wold also made an experiment with the additive case on an artificial example consisting of two tests with one common factor and the size of the sample was 50. He found that Whittle's procedures worked quite satisfactorily in all three examples and gave acceptable figures both for loadings and for individual factor values.

Whittle's approach has the advantage over Lawley's that the solutions for loadings and factor scores do not require any stipulations regarding the form of the distribution of the factor scores or test scores. On the other hand, the two different assumptions as to the size of the unique variance in relation to the test variance involve limitations that Lawley does not need to make. The fact that it is not necessary to know their numerical values, and that we need only assume that they are equal after suitable standardization of the different test variances, is, however, a relief. It would be of value, if investigations were made in the future to ascertain how strictly the requirement, that the unique variances must be equal in size, need be satisfied in order that the principal component solutions may work satisfactorily.

After analysis by Whittle's approach, confidence intervals can be obtained for loadings and factor scores and hypotheses as to the number of significant factors tested by Bartlett's test. This is based on the sampling theory developed by Wilks for the roots of a covariance matrix (Bartlett, 1950 and 1951). It assumes that the unique factors are statistics with normal distribution. It means that the significance test, contrary to the method of analysis, requires that both errors and specific factors are normally distributed. If this is assumed for the specific factors, then it would be reasonable to consider also the common factors as statistics with normal distribution. The fact is, that there is no once-and-for-all given limit or factual difference between the common and specific parts of the variance in a certain variable. Common variance for a test included in a certain battery can, if the test is analyzed in another battery, fully or partly become specific variance and contrariwise. However, it has not yet been determined how strictly the requirement of normality must be satisfied in order for the test of significance to function satisfactorily. It should also be pointed out that the assumption of normality is necessary only for the test of significance but not for the principal component solution itself.

It should be added that Bartlett and Lawley in discussions with the author of this monograph were critical of the possibility of conducting an analysis according to Whittle's methods, under the assumption that both the common factor loadings and factor scores are parameters to be estimated, since the number of parameters increases in proportion to the number of observations. Kendall (1954, p. 483) makes the same point in a critical review of the Uppsala Symposium.

> There is a substantial difficulty here because the number of parameters under estimate exceeds the number of observations.

Whittle and Wold reply to this criticism that: There are two problems of estimation involved; the first refers to the estimation of $n \times m$ common factor loadings, and the second refers to the estimation of $N \times m$ common factor scores. The errors of estimation of the loadings approach zero when the size of the sample, N, approaches infinity, but this is not true for the errors in estimating the factor scores. On the other hand, if the number of tests, n, increases, the errors in estimating the loadings will not approach zero, while the errors in estimating factor scores will approach zero.

As mentioned in Chapter I, in behavioral science it is considered more important at present to obtain estimates of the loadings than of the factor scores. It will be possible to obtain acceptable estimates of the loadings by Whittle's methods, especially if the size of the sample, N, is large and the number of common factors, m, is small.

Lawley (1953) presented his Method III, which was partly influenced by Whittle's investigations, at the Uppsala Symposium.[1] With this method he obtains estimates of the factor loadings, and makes the same two assumptions as to the unique variances as Whittle does. The method works, however, with the assumption that the factor scores are normally distributed statistical variates. The method is also provided with a test of significance for large samples. Lawley has tested his new method on an artificial sample of 400 observations, 7 tests and 2 common factors, and has obtained good results (Lawley, 1955).

Whittle's methods and Lawley's method III have certain drawbacks in common. In the additive case the unique factor loadings are assumed to be equal after suitable adaptation of the size of the different test variances. Some knowledge of the size of the unique variances is necessary for this adaptation. In the multiplicative case the residuals must form a constant proportion of the test variances, a condition which will seldom apply in practice. It is not yet clear how sensitive the methods are to biased or inefficient estimates of the unique variances or of their relative size.

Pending further investigations of the characteristics of Whittle's approach, Lawley's method I would seem at present to be the most useful, from a practical point of view, of all the methods which are firmly based on statistical principles, since it does not require any knowledge of the size of the unique variances. It presupposes, however, in principle, normal distribution of the factor scores and test scores. As has already been pointed out, however, Andersson and Rubin have indicated that the method works fairly well for large samples even when these assumptions are not satisfied.

---

[1] Lawley (1942) has also devised a Method II, which, however, has been shown not to converge towards an acceptable solution (Bartlett, 1950, pp. 80–81, and Whittle, 1952, p. 224).

Unless electronic computers are available, Lawley's method I requires a great deal of work, especially as it seems to be a necessary condition that the initial set of trial values for the factor loadings should lie very close to the solution, if the convergence is to function properly. (Lord, 1956, p. 39.) This difficulty can be overcome to some extent by the use of centroid loadings as trial values in the iteration. The centroid method is often able to produce loadings that closely correspond to maximum likelihood estimates. (Lawley, 1955).

In order to investigate this correspondence Lawley has compared his method III with a centroid method on large samples from a population with normal distribution. He selected his method III instead of I, for the following reasons (op. cit., p. 175):

> Our main justification for the assumption will, however, be that the centroid method is to be compared with a maximum likelihood method in which also the residual variances are assumed to be known. It might reasonably be expected that any inaccuracies in the values of the residual variances would have much the same effect on both methods and would not seriously invalidate the comparison.

In discussions with the author, Lawley expressed the opinion that the results of the comparisons are, on the whole, also applicable to method I. In order to facilitate the investigation, Lawley describes a modified form of the centroid solution. It has the following characteristics (op. cit., p. 176):

(i) the residual variances are assumed to be known,
(ii) the sign changes are determined in advance,
(iii) the process is applied to the covariance matrix.

Lawley considers that the results of his investigations show that the centroid method is more efficient than is generally believed. He concludes his analysis by saying (op. cit., p. 189):

> To sum up, we may say that the centroid method, at least in its modified form, has in general fairly high efficiency and has the advantage of being relatively simple to perform, as compared with the maximum likelihood method. It has the disadvantage that the χ^2 test of significance is not so convenient and exists as yet only in a large-sample form.

The twelve correlation matrices, based on twelve artificial samples which are mentioned in connection with the discussion about communalities, make it possible to compare the centroid method with Lawley's method No. I in the case where the expected rank is equal to one. In connection with another investigation (Henrysson, 1950), the factor loadings were calculated by Lawley's method I for covariance matrices based upon the same data as those relating to the twelve correlation matrices. As Lawley (1940, pp. 69–70) has shown theoretically and Burt (1950 b, pp. 161–163) empirically, the loadings according to Lawley's method I in both types of matrices stand in a simple relation to each other. Each loading in the correlation-matrix can be calculated from the loadings of the covariance matrix by dividing them by the square root of the variance of the test in question.

The centroid loadings for the twelve correlation matrices have been calculated and are given in the appendix. The communalities were calculated through iteration of the centroid solution until the communalities were equal to the squares of the centroid loadings.

The loadings according to the centroid method and Lawley's method No. I are given under each correlation matrix in the appendix. The loadings according to both methods are very close to each other. Out of the $12 \times 9 = 108$ pairs of loadings it is only 21, which differ by more than 0.002. The conclusion reached is, that at least in this simple case, both methods give loadings which are very similar.

Tests of Significance for Factor Loadings

In factor analysis there are two related problems regarding tests of significance, both of which assume a knowledge of confidence limits (Burt, 1952, pp. 131–133). The first problem is when to stop factoring. This problem assumes that there are confidence limits which make it possible to test whether the differences between the matrix $A_m \phi A_m'$ and the correlation matrix R are significant or not. Several methods for accomplishing this have already been discussed in this chapter.

The second problem is concerned with confidence limits for different loadings in matrix A_m. Knowledge of such limits should, for example, make it possible to test whether a certain estimated loading deviates from zero or not.

Psychologists have often tried to establish confidence limits for factor loadings by very rough methods. Certain rather arbitrary rules of thumb are used for the investigation of large samples, generally without adequate reasons. It is customary, for example, to consider loadings outside the interval of 0.30 to $\div 0.30$ as significant, while loadings in the interval of 0.10 to $\div 0.10$ are considered to be approximately zero.

A number of attempts have been made to find formulae for the standard error of a single factor loading, including zero loadings. These formulae are discussed by Burt (1952, p. 127–133).

The only acceptable tests of significance for loadings would seem to be those introduced by Lawley (1949) and Whittle (1952). These tests are, however, based on the same assumptions as those in respect of their methods of analysis. Factor matrices calculated by these methods must, however, be rotated to give meaningful solutions in behavioral science.

A procedure, which accepts as significant loadings outside certain limits around zero, and which considers as zero loadings those which lie within a certain interval around zero, is unsatisfactory from a statistical viewpoint, even if the limits are based on a knowledge of the sampling distribution. When dealing with a large number of loadings, this procedure can result both in many errors of acceptance and of rejection.

It would be more satisfactory if the testing of significance could be carried out in connection with hypotheses as to the whole factor structure. The investigation of significance could be conducted, for example, by studying the sum of the squares of the differences between empirical loadings and loadings in accordance with an hypothesis. Such hypotheses could be based on a certain theory or on earlier factor analyses. In the later case, the comparison is at the same time a study of invariance, as the basis for the hypothesis is a factor matrix obtained in another investigation. This matrix is compared with the factor matrix in a new investigation with the same, or with partly the same battery of tests.

It is also conceivable to test less specified hypotheses, for example, that simple structure exists in regard to a certain test battery. In this case, the hypothesis implies a number of zero loadings distributed according to the principle of simple structure. The testing of such a hypothesis, which is difficult to formalize, particularly in the oblique case, could, for example, consist in minimizing a sum of the squares of the deviations from a factor matrix containing suitably placed zero loadings. A solution could perhaps be reached, if Lawley's or Whittle's methods were further developed so that certain restrictions, for example, the presence of zero loadings, could be introduced and tested. At present, there is no general method available for testing this type of hypotheses as to factor structure, although attempts have been made by Howe (1955, pp. 104–114) and Anderson& Rubin (1956, pp. 137–138). They discuss restrictions on maximum likelihood solutions, which imply zero loadings according to modifications of the principles of simple structure. By this or similar methods the two problems of confidence limits could be merged into one, viz. the problem of finding confidence limits for the differences in

$$R - A_m \phi A'_m \tag{102}$$

with certain restrictions imposed on A_m.

SUMMARY

This monograph is devoted to an investigation of what type of problems in the behavioral sciences may be solved by factor analysis, how factorial studies ought to be designed and carried out, and how some of the technical problems in connection with the methods of analysis may be solved. The discussion is concerned with the traditional form of factor analysis, the so-called *R*-technique.

Chapter I contains a short review of some important aspects of the mathematical theory of factor analysis. The basic equations and their geometrical representation are given and discussed. It is pointed out that the general equations do not give unique solutions. Limiting restrictions on the solution must be made. In this chapter the assumptions, as to the nature of the data, are considered, which have to be satisfied if factor analysis is to give meaningful results.

Chapter II surveys the most important schools of factor analysis and the models they use. The limiting restrictions which are made when employing different models are given prominence. The main differences between the various schools refer to the principles governing the choice of the pattern of zero loadings, and to what extent negative loadings should be accepted. It is generally agreed that factor analysis is a useful descriptive tool; but some leading scientists in the field, mainly Spearman and Thurstone also wish, under certain conditions, to assign to the factors explanatory properties which refer to the nature of the underlying structure.

Chapter III discusses three main types of application of factor analysis. The first refers to the problem of finding the number of dimensions. It is pointed out that factor analysis has made an important contribution to psychological theory and application by stressing, and giving more exact meaning to the concept of dimensionality. This refers especially to the field of scaling. In this connection, the concept of homogeneity is discussed and clarified. The second type of application consists in employing factor analysis for finding condensed and simple descriptions of a large number of observed variables and for describing the relations between them. The choice of the model, in this case, is a matter of convenience, where certain general requirements in respect of the descriptive models or of the special field of application are decisive. How factor analysis can help in finding batteries consisting of valid and low-correlated tests which are useful for predictive purposes is also discussed. This is applied especially to the field of differential prediction; and it is shown that here the use of factor tests is an attractive solution from a theoretical point of view. Though the problems involved are many, the practical results obtained are few. The third type of application comprises the use of factor analysis for finding factors with explanatory properties.

In Chapter IV the distinction between descriptive and explanatory factor analysis is considered. For this purpose, results are applied which are obtained in connection with the theory of science and the discussion of intervening variables and theoretical constructs. The main finding of this investigation is, that it seems possible to distinguish between descriptive and explanatory factor analysis, even if this is difficult to apply in practice. Descriptive factor analysis is, in principle, only a deductive procedure which gives operationally defined factors related to a certain set of data. Explanatory factor analysis is an inferential procedure, the object of which is to find generalized factors with surplus content referring to regularities and relations outside an actual factor study. Invariance is an important requirement in respect of explanatory factors.

Chapter V deals with the problem of how explanatory factor analysis ought to be designed and carried out. This can be done either in a hypothesis-testing or in a hypothesis-generating manner.

Chapter VI contains a discussion of factor analysis conducted according to the principles of simple structure, and of some of the technical problems involved. The main principle is to find a set of hyperplanes of lower dimensionality than the rank of the correlation matrix. Moreover, it is shown that the principles of simple structure, as at present stated, are not always sufficient to afford unique solutions, especially in the case of rotation to an oblique solution. Different analytical methods for rotation are discussed, and it is shown that none of them as at present constituted, is fully satisfactory, and needs to be supplemented by graphical rotation. Requirements in respect of the tests in the battery are examined, and it is shown how differences in the reliability of the tests may interfere with the choice of hyperplanes during the rotational procedure. The problem of estimating communalities is treated. For all applicable factorial methods, except Lawley's method I, it is necessary to make, in advance, estimates of, or certain assumptions as to communalities. In this connection, it is demonstrated both algebraically and by experiments on artificial samples that a principal component method without communalities, as suggested by Thurstone, is useful only in the case of rank one.

Chapter VII deals with the problem of obtaining invariance between different factor solutions, after changes have been made in the composition of the test battery or in the population. Methods based on the principles of simple structure can in many cases give invariant solutions. Subjective inspection and other inefficient methods of studying invariance are criticized. Different methods for studying numerical invariance are discussed. These methods require that the same units of measurement are used for a test analyzed in different factor studies. This often implies that co-variances instead of correlations should be analyzed. It is pointed out that an attractive way of studying numerical invariance would be to consider factor patterns found in earlier factor studies as hypotheses which require to be tested by new factor analyses. Factor analytical studies of invariance ought to be supplemented by non-factorial validation of the factors.

In Chapter VIII factor analysis is considered as a technique of statistical inference. Many of the methods used by psychologists are unsatisfactory from a statistical point of view and lack acceptable tests of significance. Methods developed by statisticians are generally based on maximum-likelihood or on principal-component solutions. These methods are discussed mainly from the point of view of the underlying assumptions and of their applicability to the behavioral sciences. It is pointed out that these methods should be adjusted so as to give solutions which are meaningful for data in the behavioral sciences. The available statistical methods for testing hypotheses as to invariance are still not satisfactory. Current statistical methods for testing significance should be adjusted so as to make it possible to test hypotheses as to explanatory factors and their invariance.

APPENDIX

Artificial correlation matrices and loadings found by four different methods

Row 1 under each correlation matrix contains loadings by the principal component method with communalities and row 2 without communalities. Row 3 contains loadings by Lawley's method I and row 4 loadings by the centroid method. The chi-square value under row 4 is calculated from the solution by Lawley's method I.

SAMPLE I

Variable	1	2	3	4	5	6	7	8	9
1		.471	.546	.499	.483	.521	.510	.530	.537
2	.471		.459	.434	.455	.474	.419	.452	.432
3	.546	.459		.567	.444	.525	.481	.592	.504
4	.499	.434	.567		.528	.512	.491	.493	.500
5	.483	.455	.444	.528		.490	.475	.473	.552
6	.521	.474	.525	.512	.490		.541	.518	.508
7	.510	.419	.481	.491	.475	.541		.439	.542
8	.530	.452	.592	.493	.473	.518	.439		.516
9	.537	.432	.504	.500	.552	.508	.542	.516	
PC									
with h^2	.729	.628	.734	.714	.688	.726	.689	.713	.727
without h^2	.729	.629	.734	.714	.688	.726	.689	.713	.727
Lawley	.729	.628	.735	.714	.687	.725	.690	.714	.727
Centroid	.729	.629	.733	.714	.689	.727	.689	.712	.728

$\chi^2 = 21.8$

SAMPLE II

Variable	1	2	3	4	5	6	7	8	9
1		.644	.487	.601	.616	.555	.548	.637	.527
2	.644		.544	.583	.622	.485	.589	.546	.573
3	.487	.544		.488	.599	.580	.521	.566	.540
4	.601	.583	.488		.604	.529	.619	.585	.562
5	.616	.622	.599	.604		.559	.558	.619	.611
6	.555	.485	.580	.529	.559		.554	.607	.517
7	.548	.589	.521	.619	.558	.554		.548	.567
8	.637	.546	.566	.585	.619	.607	.548		.583
9	.527	.573	.540	.562	.611	.517	.567	.583	
PC									
with h^2	.768	.761	.711	.759	.799	.723	.745	.781	.741
without h^2	.768	.761	.711	.759	.799	.723	.745	.781	.741
Lawley	.771	.763	.711	.760	.799	.721	.743	.779	.741
Centroid	.767	.761	.712	.758	.799	.723	.745	.781	.741

$\chi^2 = 38.6$

SAMPLE III

Variable	1	2	3	4	5	6	7	8	9
1		.510	.540	510	.549	.521	.534	.535	.490
2	.510		.528	.426	.442	.452	.414	.531	.419
3	.540	.528		.582	.532	.529	.426	.526	.495
4	.510	.426	.582		.470	.552	.464	.505	.489
5	.549	.442	.532	.470		.536	.428	.503	.478
6	.521	.452	.529	.552	.536		.464	.520	.535
7	.534	.414	.426	.464	.428	.464		.404	.482
8	.535	.531	.526	.505	.503	.520	.404		.544
9	.490	.419	.495	.489	.478	.535	.482	.544	
PC									
with h^2	.747	.656	.743	.711	.699	.732	.634	.725	.696
without h^2	.747	.655	.743	.711	.699	.733	.633	.724	.696
Lawley	.744	.657	.743	.713	.701	.732	.633	.724	.695
Centroid	.749	.655	.743	.710	.698	.733	.633	.724	.696

$\chi^2 = 24.2$

SAMPLE IV

Variable	1	2	3	4	5	6	7	8	9
1		.521	.478	.516	.513	.440	.523	.494	.462
2	.521		.483	.483	.530	.439	.502	.499	.503
3	.478	.483		.523	.486	.453	.543	.476	.541
4	.516	.483	.523		.497	.532	.529	.544	.476
5	.513	.530	.486	.497		.512	.457	.559	.503
6	.440	.439	.453	.532	.512		.465	.469	.444
7	.523	.502	.543	.529	.457	.465		.482	.490
8	.494	.499	.476	.544	.559	.469	.482		.505
9	.462	.503	.541	.476	.503	.444	.490	.505	
PC									
with h^2	.700	.702	.707	.730	.721	.661	.708	.716	.695
without h^2	.700	.702	.707	.730	.721	.662	.708	.716	.695
Lawley	.701	.702	.706	.729	.720	.663	.708	.717	.695
Centroid	.700	.702	.707	.731	.722	.661	.708	.716	.695

$\chi^2 = 19.2$

SAMPLE V

Variable	1	2	3	4	5	6	7	8	9
1		.543	.388	.398	.452	.489	.488	.473	.490
2	.543		.512	.513	.587	.577	.578	.482	.542
3	.388	.512		.516	.512	.503	.500	.435	.492
4	.398	.513	.516		.401	.503	.515	.426	.434
5	.452	.587	.512	.401		.516	.506	.422	.485
6	.489	.577	.503	.503	.516		.461	.512	.484
7	.488	.578	.500	.515	.506	.461		.444	.517
8	.473	.482	.435	.426	.422	.512	.444		.527
9	.490	.542	.492	.434	.485	.484	.517	.527	
PC									
with h^2	.661	.787	.687	.657	.695	.725	.718	.659	.710
without h^2	.661	.786	.687	.657	.695	.725	.718	.659	.709
Lawley	.663	.788	.685	.658	.699	.723	.718	.656	.707
Centroid	.660	.786	.687	.657	.692	.726	.719	.660	.711

$\chi^2 = 28.1$

SAMPLE VI

Variable	1	2	3	4	5	6	7	8	9
1		.422	.501	.372	.423	.456	.464	.413	.396
2	.422		.540	.477	.481	.522	.511	.431	.459
3	.501	.540		.517	.490	.561	.575	.476	.524
4	.372	.477	.517		.487	.538	.558	.468	.518
5	.423	.481	.490	.487		.490	.441	.497	.468
6	.456	.522	.561	.538	.490		.577	.456	.575
7	.464	.511	.575	.558	.441	.577		.450	.531
8	.413	.431	.476	.468	.497	.456	.450		.478
9	.396	.459	.524	.518	.468	.575	.531	.478	
PC									
with h^2	.607	.687	.756	.706	.671	.756	.742	.650	.709
without h^2	.607	.687	.756	.707	.670	.756	.742	.649	.710
Lawley	.608	.687	.755	.709	.666	.757	.745	.645	.712
Centroid	.607	.687	.757	.705	.673	.755	.741	.651	.708

$\chi^2 = 17.8$

SAMPLE VII

Variable	1	2	3	4	5	6	7	8	9
1		.591	.542	.469	.438	.580	.541	.565	.549
2	.591		.578	.573	.583	.616	.548	.553	.551
3	.542	.578		.501	.513	.454	.458	.530	.513
4	.469	.573	.501		.515	.583	.522	.515	.564
5	.438	.583	.513	.515		.519	.495	.500	.476
6	.580	.616	.454	.583	.519		.595	.523	.494
7	.541	.548	.458	.522	.495	.595		.516	.486
8	.565	.553	.530	.515	.500	.523	.516		.507
9	.549	.551	.513	.564	.476	.494	.486	.507	
PC									
with h^2	.736	.798	.698	.728	.689	.753	.713	.721	.708
without h^2	.736	.799	.699	.728	.690	.754	.713	.722	.708
Lawley	.737	.798	.698	.728	.691	.756	.714	.720	.707
Centroid	.736	.799	.699	.729	.689	.753	.713	.722	.709

$\chi^2 = 31.6$

SAMPLE VIII

Variable	1	2	3	4	5	6	7	8	9
1		.615	.559	.576	.544	.446	.551	.557	.553
2	.615		.509	.580	.561	.467	.540	.517	.522
3	.559	.509		.540	.463	.487	.525	.507	.555
4	.576	.580	.540		.631	.560	.595	.574	.618
5	.544	.561	.463	.631		.545	.555	.553	.533
6	.446	.467	.487	.560	.545		.541	.548	.519
7	.551	.540	.525	.595	.555	.541		.566	.662
8	.557	.517	.507	.574	.553	.548	.566		.547
9	.553	.522	.555	.618	.533	.519	.662	.547	
PC									
with h^2	.743	.725	.695	.797	.741	.689	.770	.737	.765
without h^2	.743	.726	.694	.797	.741	.689	.770	.737	.765
Lawley	.742	.726	.694	.796	.742	.689	.771	.735	.767
Centroid	.743	.726	.695	.797	.741	.689	.770	.737	.765

$\chi^2 = 31.9$

SAMPLE IX

Variable	1	2	3	4	5	6	7	8	9
1		.540	.574	.566	.448	.420	.513	.512	.437
2	.540		.483	.504	.472	.406	.500	.466	.432
3	.574	.483		.479	.445	.533	.491	.470	.529
4	.566	.504	.479		.459	.483	.449	.488	.505
5	.448	.472	.445	.459		.424	.515	.467	.486
6	.420	.406	.533	.483	.424		.445	.417	.430
7	.513	.500	.491	.449	.515	.445		.487	.483
8	.512	.466	.470	.488	.467	.417	.487		.522
9	.437	.432	.529	.505	.486	.430	.483	.522	
PC									
with h^2	.730	.686	.727	.712	.667	.636	.701	.691	.689
without h^2	.730	.686	.726	.712	.666	.636	.701	.690	.689
Lawley	.733	.688	.726	.711	.665	.636	.699	.690	.688
Centroid	.728	.685	.727	.712	.667	.635	.701	.691	.689

$\chi^2 = 28.2$

SAMPLE X

Variable	1	2	3	4	5	6	7	8	9
1		.491	.430	.407	.480	.403	.352	.479	.434
2	.491		.519	.542	.464	.487	.457	.543	.450
3	.430	.519		.505	.441	.442	.415	.459	.520
4	.407	.542	.505		.485	.415	.463	.531	.475
5	.480	.464	.441	.485		.454	.480	.477	.430
6	.403	.487	.442	.415	.454		.459	.456	.431
7	.352	.457	.415	.463	.480	.459		.550	.430
8	.479	.543	.459	.531	.477	.456	.550		.490
9	.434	.450	.520	.475	.430	.431	.430	.490	
PC									
with h^2	.632	.731	.684	.704	.678	.645	.659	.738	.668
without h^2	.632	.731	.684	.705	.678	.645	.659	.737	.669
Lawley	.633	.731	.683	.708	.675	.644	.661	.737	.667
Centroid	.631	.731	.684	.703	.680	.646	.658	.738	.669

$\chi^2 = 22.1$

SAMPLE XI

Variable	1	2	3	4	5	6	7	8	9
1		.568	.516	.520	.527	.584	.530	.505	.504
2	.568		.468	.511	.502	.566	.475	.438	.466
3	.516	.468		.500	.510	.492	.512	.499	.518
4	.520	.511	.500		.572	.573	.526	.527	.594
5	.527	.502	.510	.572		.560	.513	.501	.535
6	.584	.566	.492	.573	.560		.558	.462	.552
7	.530	.475	.512	.526	.513	.558		.525	.621
8	.505	.438	.499	.527	.501	.462	.525		.485
9	.504	.466	.518	.594	.535	.552	.621	.485	
PC									
with h^2	.736	.687	.690	.751	.730	.757	.739	.676	.742
without h^2	.736	.687	.690	.751	.730	.757	.738	.676	.742
Lawley	.734	.688	.688	.751	.730	.757	.739	.675	.744
Centroid	.737	.686	.690	.751	.730	.756	.739	.676	.742

$\chi^2 = 25.8$

SAMPLE XII

Variable	1	2	3	4	5	6	7	8	9
1		.485	.500	.486	.468	.556	.538	.483	.501
2	.485		.445	.400	.420	.510	.428	.473	.424
3	.500	.445		.435	.502	.458	.524	.507	.500
4	.486	.400	.435		.485	.472	.487	.440	.471
5	.468	.420	.502	.485		.561	.494	.513	.465
6	.556	.510	.458	.472	.561		.502	.536	.512
7	.538	.428	.524	.487	.494	.502		.533	.496
8	.483	.473	.507	.440	.513	.536	.533		.518
9	.501	.424	.500	.471	.465	.512	.496	.518	
PC									
with h^2	.722	.635	.692	.653	.700	.741	.720	.720	.696
without h^2	.722	.636	.692	.653	.701	.741	.720	.720	.696
Lawley	.721	.637	.691	.654	.701	.740	.720	.720	.696
Centroid	.722	.635	.692	.653	.700	.741	.720	.720	.696

$\chi^2 = 15.5$

REFERENCES

AHMAVAARA, YRJÖ (1954a) The mathematical theory of factorial invariance under selection. *Psychometrika*, 19, 27–38.

—— (1954b) *Transformation Analysis of Factorial Data*. Annales Academiae Scientiarum Fennicae, Ser. B. 88, 2.

—— (1957) *On the Unified Factor Theory of Mind*. Annales Academiae Scientiarum Fennicae, Ser. B. 106.

ALLPORT, GORDON, W. (1945) *Personality. A Psychological Interpretation*. New York: Henry Holt & Co.

ALLPORT, GORDON, W. & ODBERT, H. S. (1936) *Trait-Names: A Psycho-lexical Study*. Psychol. Monogr., No. 211.

ANDERSON, T. W. & RUBIN, H. (1956) Statistical inference in factor analysis. *Proc. Third Berkeley Symposium on Mathematical Statistics and Probability. Vol. 5*. Univ. Calif. Press, 110–150.

BARGMANN, R. (1951) *Identification of Simple Structure by Successive Extension*. Research Report. Chicago: Univ. Chicago.

—— (1953) *A Comparison of New Analytical Methods for the Determination of Simple Structure* Hochschule f. Int. Päd. Forschung Frankfurt/Main, Sept. 1953.

—— (1955) *A Demonstration Study on the Effectiveness of Factor Analytical Methods*. Hochschule f. Int. Päd. Forschung Frankfurt/Main, June 1955.

BARTLETT, M. S. (1950) Tests of significance in factor analysis. *Brit. J. Psychol. Stat. Sect.*, 3, 77–85.

—— (1951) A further note on tests of significance in factor analysis. *Brit. J. Psychol. Stat. Sect.*, 4, 1–2.

BERGMANN, GUSTAV (1953) Theoretical psychology. *Annual Rev. of Psychol.*, 4, 435–458.

BOND, GUY L. & CLYMER, THEODORE W. (1955) Interrelationship of the SRA primary mental abilities, other mental characteristics, and reading ability. *Journal of Educational Research*, 49, 131–141.

BRAITHWAITE, RICHARD BEVAN (1953) *Scientific Explanation*. Cambridge, at the Univ. Press.

BROWN, W. & STEPHENSON, W. (1933) A test of the theory of two factors. *Brit. J. Psychol.*, 23, 352–370.

BURT, CYRIL (1940) *The Factors of the Mind*. London: Univ. London Press.

—— (1949a) The two-factor theory. *Brit. Journ. Psychol. Stat. Sect.*, 2, 151–179.

BURT, CYRIL (1949b) Subdivided factors. *Brit. Journ. Psychol. Stat. Sect.*, 2, 41–63.

—— (1950a) Group factor analysis. *Brit. J. Psychol. Stat. Sect.*, 3, 40–76.

—— (1950b) Maximum likelihood loadings as obtained from single-factor correlation tables. *Brit. J. Psychol. Stat. Sect.*, 3, 161–165.

—— (1951) The factorial study of the mind. *Essays in Psychology Dedicated to David Katz*. Uppsala: Almqvist & Wiksell.

—— (1952) Tests of significance in factor studies. *Brit. J. Psychol. Stat. Sect.*, 5, 109–133.

—— (1953) Scale analysis and factor analysis. *Brit. Journ. Stat. Psychol.*, 6, 5–23.

CARLSSON, N. GÖSTA (1949) *Dimensions of Behaviour*. Lund: Gleerup.

CARNAP, RUDOLF (1950) *Locigal Foundations of Probability*. Chicago: Univ. Chicago Press.

—— (1956) The methodological character of theoretical concepts. *Minnesota Studies in the Philosophy of Science*. 1, 38–76.

CARROLL, J. B. (1953) An analytical solution for approximating simple structure in factor analysis. *Psychometrika*, 18, 23–38.

CATTELL, RAYMOND B. (1946) *Description and Measurement of Personality*. New York: World Book Company.

—— (1952) *Factor Analysis*. New York: Harper.

CATTELL, R. B. & BAGGALEY, A. R. (1954) The salient variable similarity index-s-for factor matching,. *Advance Publ. Lab. Pers. Assessment* & *Group Behavior. Univ. Illinois*. No. 4.

CATTELL, RAYMOND B. & CATTELL, A. K. S. (1955) Factor rotation for proportional profiles. *Brit. J. Stat. Psychol.*, 8, 83–93.

COOMBS, C. H. (1952) A theory of psychological scaling. *Engineering Res. Bull.* Ann Arbor: Univ. Michigan. No. 34.

COOMBS, C. H. & SATTER, G. A. (1949) A factorial approach to job families. *Psychometrika*, 14, 33–42.

CRONBACH, LEE J. (1956) Assessment of individual differences. *Annual Rev. of Psychology*, 7, 173–196.

CRONBACH, LEE J. & MEEHL, P. E. (1955) Construct validity in psychological tests. *Psychol. Bull.*, 52, 281–302.

CURETON, EDWARD E. (1951) Validity. *In* LINDQUIST E. F. 1951.

DAILEY, JOHN T. (1951) The supply and identification of high level Talent. *Proc. Invit. Conf. Testing Problems*. Princeton: Educ. Testing Service, 17–23.

DWYER, PAUL S. (1939) The contribution of an orthogonal multiple factor solution to multiple correlation. *Psychometrika*, 4, 163–171.

—— (1940) The evaluation of multiple and partial correlation coefficients from the factorial matrix. *Psychometrika*, 5, 211–232.

—— (1954) Solution of the personnel classification problem with the Method of optimal regions. *Psychometrika*, 18, 11–26.

Ekman, G. (1954) Dimensions of color vision. *J. Psychol.*, 38, 467–474.

Elmgren, John (1955) *Psykologisk faktoranalys*. Stockholm: Natur & Kultur.

Emmett, W. G. (1936) Sampling error and the two-factor theory. *Brit. J. Psychol.*, 26, 362–387.

Eysenck, H. J. (1953) The logical basis of factor analysis. *Amer. Psychologist*, 8, 105–114.

—— (1954) *The Psychology of Politics*. London: Routledge & Kegan Paul.

Feigl, H. (1953) Some remarks on the meaning of scientific explanation. *Readings in Philosophical Analysis*. Edit. Feigl, H. & Sellars, W. New York, Appleton-Century-Crofts, 510–514.

Ferguson, George A. (1954) The concept of parsimony in factor analysis. *Psychometrika*, 19, 281–290.

Fleishman, Edvin A. (1953) A factor analysis of intra-task performance on two psychomotor tests. *Psychometrika*, 18, 45–55.

Fleishman, Edvin H. & Hempel, Walter E. (1954) Changes in factor structure of a complex psychomotor test as a function of practice. *Psychometrika*, 19, 239–252.

French, J. W. (1951) The description of aptitude and achievement tests in terms of rotated factors. *Psychometric Monogr.* no. 5, Chicago: Univ. Chicago Press.

—— (1954) *Kit of Selected Tests for Reference Aptitude and Achievement Factors*. Princeton: Educational Testing Service.

—— (1955) Comparative prediction and other techniques for use with a guidance battery. *Research Bulletin* 55–4. Princeton: Educational Testing Service.

Fruchter, Benjamin (1954) *Introduction to Factor Analysis*. New York: D. van Nostrand Co.

Ghiselli, Edwin E. (1956) Dimensional problems of criteria. *J. Appl. Psych.* 40, 1–4.

Ginsberg, Arthur (1954) Hypothetical constructs and intervening variables. *Psychol. Review*, 61, 119–131.

Gourlay, Neil (1951) Difficulty factors arising from the use of tetrachoric correlations in factor analysis. *Brit. J. Psychol. Stat. Sect*, 4, 65–76.

Gragg, Donald B. & Gordon, Mary Agnes (1950) *Validity of the Airman Classification Battery AC-1*. Research Bull 50–3. Air Training Command. Human Resources Research Center, Lackland Air Force Base.

Guilford, J. P. (1948) Factor analysis in a test development program. *Psychol. Rev.*, 55, 79–94.

—— (1954) *Psychometric Methods*. Second Edition. New York: McGraw-Hill.

Guilford, J. P. & Michael, W. B. (1950) Changes in common-factor loadings as tests are altered homogeneously in length. *Psychometrika*, 15, 237–251.

Gulliksen, H. (1950) *Theory of Mental Tests.* New York: Wiley.

Guttmann, Louis (1940) Multiple rectilinear prediction and the resolution into components. *Psychometrika*, 5, 75–99.

—— (1950) The basis for scalogram analysis. *In* Stouffer *et al.* 1950.

Hartley, Raymond E. (1952) *On the Logical Foundation of Factor Analysis.* Thesis No. 1638, Chicago: Univ. Chicago. Library Department of Photographic Reproduction.

—— (1954) Two kinds of factor analysis. *Psychometrika*, 19, 195–203

Hempel, Carl G. (1952) Fundamentals of concept formation in empirical science. *Intern. Encyclopedia of Unified Science.* Vol. 2, No. 7. Chicago: Univ. of Chicago Press.

Henrysson, Sten (1950) The significance of factor loadings. *Brit. J. Psychol. Stat. Sect.*, 3, 159–165.

Holzinger, Karl & Harman, Harry H. (1941) *Factor Analysis.* Chicago: Univ. Chicago Press.

Holzinger, Karl J. & Swineford, Frances (1939) *A Study in Factor Analysis: The Stability of a Bi-factor Solution.* Supplementary Educational Monographs. No. 48. Chicago: The Univ. of Chicago.

Horst, Paul *et al.* (1941) *The Prediction of Personal Adjustment.* New York: Social Science Research Council.

Hotelling, H. (1935) The most predictable criterion. *Journ. Educ. Psychol.*, 26, 139–142.

—— (1942) Review of Holzinger & Harman: Factor Analysis. *Science*, 95, 504–507.

Howe, William G. (1955) *Some Contributions to Factor Analysis.* ORNL-1919, Physics. Oak Ridge National Laboratory.

Hull, C. L. (1943) *Principles of Behaviour.* New York: Appleton-Century Co.

Kendall, M. G. (1950) Factor analysis as a statistical technique. *Journ. Roy. Stat. Soc.* (*Series B*), 12, 60–73.

—— (1954) Review of the Uppsala Symposium. *J. Roy. Stat. Soc.* (*Series B*), 16, 482–484.

Kerschner, Allan M. (1955) *A Report on Job Analysis.* ONR Report ACR-5. Washington D.C.: Office of Naval Research. Department of the Navy.

Lawley, D. N. (1940) The estimation of factor loadings by the method of maximum likelihood. *Proc. Roy. Soc. Edinb.*, 60, 64–82.

—— (1942) Further investigations in factor estimation. *Proc. Roy. Soc. Edinb. Sect. A*, 61, Part 2 (No. 14), 176–185.

—— (1949) Problems in factor analysis. *Proc. Roy. Soc. Edinb. Sect. A*, 62, 394–399.

Lawley, D. N. (1953) A modified method of estimation in factor analysis and some large sample results. *Uppsala Symposium on Psychological Factor Analysis, 1953.*

—— (1955) A statistical examination of the centroid method. *Proc. Roy. Soc. Edinb. Sect. A*, 64, Part 2 (No. 13), 175–189.

Lawley, D. N. & Swanson, Z. (1954) Tests of significance in a factor analysis of artificial data. *Brit. J. Stat. Psychol.*, 7, 75–79.

Lazarsfeld, Paul F. (1950) The logical and mathematical foundation of latent structure analysis. *In* Stouffer *et al.* 1950.

Leiman, J. M. (1952) The use of common factor loadings in a multiple regression system. *Amer. Psychologist*, 7, 297.

Lindquist, E. F. *Ed.* (1951) *Educational Measurement*. Washington D.C.: Amer. Council of Educ.

Loevinger, Jane (1947) A systematic approach to the construction and evaluation of tests of ability. *Psychol. Monogr.*, 56, No. 6.

Lord, Frederic M. (1956) A study of speed factors in tests and academic grades. *Psychometrika*, 21, 31–49.

Lubin, A. & Summerfield, A. (1951) A square root method of selecting a minimum set of variables in multiple regression: I. The method. *Psychometrika*, 16, 271–284.

MacCorquodale, K. & Meehl, P. E. (1948) On a distinction between hypothetical constructs and intervening variables. *Psychol. Rev.*, 55, 95–107.

McQuitty, L. L. *et al.* (1954) An approach to isolating dimensions of job success. *J. Appl. Psychol.*, 38, 227–232.

Marx, Melvin H. (1951) Intervening variable or hypothetical construct? *Psychol. Review*, 58, 235–247.

Maze, J. R. (1954) Do intervening variables intervene? *Psychol. Review*, 61, 226–234.

Mollenkopf, William, G. (1950) Predicted differences and differences between predictions. *Psychometrika*, 15, 409–417.

Mosier, C. I. (1939) Influence of chance error on simple structure: an empirical investigation on the effect of chance error and estimated communalities on simple structure in factorial analysis. *Psychometrika*, 4, 33–44.

—— (1951 a) Problems and designs of cross-validation. *Educ. Psychol. Measurem.* 11, 5–11.

—— (1951 b) Batteries and profiles. *In* Lindquist, E. F. 1951.

Moursey, E. M. (1952) The hierarchical organization of cognitive levels. *Brit. J. Psychol. Stat. Sect.*, 5, 151–180.

Neyhaus, J. O. & Wrigley, C. (1954) The quartimax method: an analytic approach to orthogonal simple structure. *Brit. J. Stat. Psychol.*, 7, 81–91.

NORTHROP, F. S. C. (1947) *The Logic of the Sciences and the Humanities.* New York: MacMillan.

PEEL, E. A. (1948) Prediction of a complex criterion and battery reliability. *Brit. J. Psychol. Stat. Sect.*, 1, 84–94.

—— (1953) Factorial analysis as a psychological technique. *Uppsala Symposium on Psychological Factor Analysis 1953.*

PINZKA, C. & SAUNDERS, D. R. (1954) Analytic rotation to simple structure, II: extension to a oblique solution. *Research Bulletin* 54–31. Princeton: Educ. Testing Service.

RAO, C. R. (1955) Estimation and tests of significance in factor analysis. *Psychometrika*, 20, 93–111.

RASCH, G. (1953) On simultaneous factor analysis in several populations. *Uppsala Symposium on Psychological Factor Analysis 1953.*

REIERSÖL, O. (1950) On the identifiability of parameters in Thurstone's multiple factor analysis. *Psychometrika*, 15, 121–149.

REYBURN, H. A. & RAATH, M. J. (1949) Simple structure. A critical examination. *Brit. J. Psych. Stat. Sect.*, 2, 125–131.

ROSNER, B. (1948) An algebraic solution for the communalities. *Psychometrika*, 13, 181–184.

ROZEBOOM (1956) Mediation variables in scientific theory. *Psychol. Review*, 63, 249–264.

SAUNDERS, DAVID R. (1953) An analytic method for rotation to orthogonal simple structure. *Research Bulletin* 53–10. Princeton: Educational Testing Service.

SLUCKIN, W. (1956) Combining criteria of success. *Occup. Psychol.*, 30, 20–26 and 57–68.

SPEARMAN, C. (1904) General intelligence objectively determined and measured. *Amer. J. Psychol.*, 15, 201–293.

—— (1927) *The Abilities of Man.* London: MacMillan & Co.

—— (1931) What the theory of factors is not. *J. Educ. Psychol.*, 22, 112–117.

—— (1933) The factor theory and its troubles: III. Misrepresentation of the theory. *J. Educ. Psychol.*, 24, 591–601.

SPEARMAN, C. & JONES, L. L. WYNN (1950) *Human Ability.* London: MacMillan.

STOUFFER, SAMUEL A. *et al.* (1950) *Measurement and Prediction.* Princeton: Princeton Univ. Press.

SWINEFORD, FRANCES (1949) A number factor. *J. Educ. Psychol.*, 40, 157–167.

TERMAN, LEWIS M. (1954) The discovery and encouraging of exceptional talent. *Amer. Psychologist*, 9, 221–230.

THOMSON, G. H. (1933) Note on Dr William Brown's paper on a central intellective factor. *Brit. J. Psychol.*, 23, 404.

THOMSON, G. H. (1948) *Factorial Analysis of Human Ability*. London: Univ. London Press.

THURSTONE, L. L. (1947) *Multiple Factor Analysis*. Chicago: Univ. Chicago Press.

—— (1954a) An analytical method for simple structure. *Psychometrika*, 19, 173–182.

—— (1954b) A method of factoring without communalities. *Proc. Inv. Conf. Testing Problems*. Princeton: Educational Testing Service, 59–62.

TOLMAN, EDVARD C. (1932) *Purposive Behavior in Animals and Men*. New York: Century Co.

—— (1936) Operational behaviorism and current trends in psychology. *Proc. 25th Anniv. Celebr. Inaug. Grad. Stud.* Los Angeles: Univ. South. Calif. Press.

—— (1951) A psychological model. *Toward a General Theory of Action*. Harvard: Harvard Univ. Press 279–361.

TUCKER, L. R. (1951) A method of synthesis of factor analysis studies. *Personnel Research Section Research Report* No. 984.

—— (1955) The objective definition of simple structure in linear factor analysis. *Psychometrika*, 20, 209–225.

—— (1956) *An Inter-battery Method of Factor Analysis*. Princeton: Educational Testing Service.

UPPSALA SYMPOSIUM (1953) *Uppsala Symposium on Psychological Factor Analysis, 17–19 March 1953*. Nordisk Psykologi's Monogr. Series No. 3.

VERNON, P. E. (1950a) *The Structure of Human Abilities*. London: Methuen.

—— (1950b) An application of factorial analysis to the study of test items. *Brit. J. Psych. Stat. Sect.*, 3, 1–15.

VOTAW, T. F. & DAILEY, J. T. (1952) Assignment of personnel to jobs. *Research Bull.*, Air Training Command. Human Resources Research Center. Lackland Air Forces Base. 52–24.

WHITTLE, P. (1952) On principal components and least square methods of factor analysis. *Skandinavisk Aktuarietidskrift*, 1952, 223–239.

WOLD, HERMAN (1948) Random normal deviates. *Tracts for Computers*, 25. Cambridge: University Press.

—— (1953) Some artificial experiments in factor analysis. *Uppsala Symposium on Psychological Factor Analysis, 1953*.

WOLFLE, DAEL (1940) Factor Analysis to 1940. *Psychometric Monogr.* No. 3. Chicago: Univ. Chicago Press.

YOUNG, G. (1941) Maximum likehood estimation and factor analysis. *Psychometrika*, 6, 49–53.

311
HEN

311
HEN

Stockholm Studies in Educational Psycholo

PUBLISHED BY THE UNIVERSITY OF STOCKHOLM

Editor: T. HUSÉN

The series includes theses and other studies by members of the University. Subscriptions to the series and orders for single volumes should be addressed to any international bookseller or directly to the publishers Almqvist & Wiksell, Gamla Brogatan 26, Stockholm, Sweden. Universities, libraries, learned societies, and publishers of learned periodicals may obtain the volumes of the series and other publications of the University of Stockholm in exchange for their own publications. Inquiries should be addressed to Stockholms Högskolas Bibliotek, Postbox 6080, Stockholm 6, Sweden.

77

1. STEN HENRYSSON, Applicability of Factor Analysis in the Behavioral Sciences. A Methodological Study. Stockholm 1957. Price Sw. kr. 15:—.

nqvist & W ckeri AB 576055